

109652

From: Chan, Christina
Sent: Wednesday, December 03, 2003 1:57 PM
To: Davis, Minh-Tam; STIC-Biotech/ChemLib
Subject: RE: Rush search request for 09/674237

RECEIVED

DEC -3 2003

STIC-BIOTECH/CHMLIB
 (STIC)

Pl ase rush: Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
 308-3973
 CM-1, 9B19

-----Original Message-----

From: Davis, Minh-Tam
Sent: Wednesday, December 03, 2003 1:54 PM
To: Chan, Christina
Subject: Rush search request for 09/674237

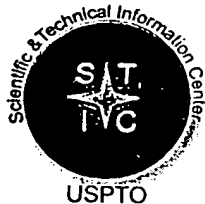
Please search in commercial database, issued patent files, PGPUB and interference:
 SEQ ID NO:1, 2, and 3
 Thank you.
 MINH TAM DAVIS
 ART UNIT 1642, ROOM 8A01, MB 8E12
 305-2008

Searcher: _____
 Phone: _____
 Location: _____
 Date Picked Up: _____
 Date Completed: _____
 Searcher Prep/Review: _____
 Clerical: _____
 Online time: _____

TYPE OF SEARCH:
 NA Sequences: _____
 AA Sequences: _____
 Structures: _____
 Bibliographic: _____
 Litigation: _____
 Full text: _____
 Patent Family: _____
 Other: _____

VENDOR/COST (where applic.)
 STN: _____
 DIALOG: _____
 Questel/Orbit: _____
 DRLink: _____
 Lexis/Nexis: _____
 Sequence Sys.: _____
 WWW/Internet: _____
 Other (specify): _____

THIS PAGE BLANK (USPTO)



STIC SEARCH RESULTS FEEDBACK FORM

Biotech-Chem Library

Questions about the scope or the results of the search? Contact *the searcher* or *contact:*

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➤ I am an examiner in Workgroup: Example: 1610

➤ Relevant prior art **found**, search results used as follows:

- ☐ 102 rejection
- ☐ 103 rejection
- ☐ Cited as being of interest.
- ☐ Helped examiner better understand the invention.
- ☐ Helped examiner better understand the state of the art in their technology.

Types of relevant prior art found:

- ☐ Foreign Patent(s)
- ☐ Non-Patent Literature
(journal articles, conference proceedings, new product announcements etc.)

➤ Relevant prior art **not found**:

- ☐ Results verified the lack of relevant prior art (helped determine patentability).
- ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 - Circ. Desk



THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:21:40 ; Search time 8605.82 Seconds
(without alignments)
17313.032 Million cell updates/sec

Title: US-09-674-237a-2
Perfect score: 3642
Sequence: 1 atgctcagttccaccacc.....tgaccaccagcgaatga 3642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBankl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hngo_hum:*
40: em_hngo_mus:*
41: em_hngo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3640.4	100.0	3723	10	AF132478 Mus muscu
2	3639.4	99.9	5145	10	AF132481 Mus muscu
3	3229.8	88.7	4025	10	AF127798 Rattus no
4	2831	77.7	5381	9	AF114488 Homo sapi
5	2830	77.7	6439	9	AF114487 Homo sapi
6	2827.8	77.6	5287	9	AF064243 Homo sapi
7	2826.8	77.6	7247	9	AF064244 Homo sapi
8	2816.2	77.3	3812	10	AF132672 Rattus no
9	2027.6	55.7	4103	5	AF032118 Xenopus l
10	1483.8	40.7	2131	6	BD127168 Primer fo
11	1483.8	40.7	2131	9	AK074554 Homo sapi
12	1259.2	34.6	2131	6	BD158570 Homo sapi
13	1259.2	34.6	2131	9	AK027846 Homo sapi
14	1104.2	30.3	1676	6	BD127640 Primer fo
15	1104.2	30.3	1676	9	AK075290 Homo sapi
16	986.4	27.1	3241	9	HSU61166 Human SH3 d
17	836.8	23.0	1299	9	BC039036 Homo sapi
18	749	20.6	1133	10	AF169621 Mus muscu
19	721	19.8	5828	6	BD167848 Method fo
20	721	19.8	5828	9	AF248540 Homo sapi
21	721	19.8	5938	9	AB033082 Homo sapi
22	713.8	19.6	4557	9	AF182199 Homo sapi
23	713.8	19.6	6091	9	AF182198 Homo sapi
24	677	18.6	4977	10	AF132479 Mus muscu
25	673.8	18.5	3594	10	AF132480 Mus muscu
26	648	17.8	4447	6	AX428899 Sequence
27	617.4	17.0	5804	10	AK122480 Mus muscu
28	531.4	14.6	78190	2	AC106760 Homo sapi
29	531.4	14.6	179264	2	AC023112 Homo sapi
30	513.8	14.1	677	6	AX677983 Sequence
31	485.4	13.3	877	6	BD124639 Primer fo
32	485.4	13.3	877	6	BD126208 Primer fo
33	477.4	13.1	2017	6	AR238090 Sequence
34	477.4	13.1	2017	6	AR257631 Sequence
35	477.4	13.1	2017	6	AR283677 Sequence
36	477.4	13.1	2017	6	AX36305 Sequence
37	435	11.9	475	10	AY127576 Mus muscu
38	387.4	10.6	1329	6	BD157252 Primer fo
39	387.4	10.6	1329	9	AK021545 Homo sapi
40	372	10.2	493	10	AF525079 Mus muscu
41	360.8	9.9	2085	9	BC038963 Homo sapi
42	342.2	9.4	1996	9	AF180522 Homo sapi
43	332.4	9.1	2873	6	AR175271 Sequence
44	331.4	9.1	1391	9	AF001630 Homo sapi
45	331.4	9.1	2004	9	AF038189 Homo sapi

ALIGNMENTS

RESULT 1
AF132478
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999
DEFINITION Mus musculus Ebel protein mRNA, complete cds.
ACCESSION AF132478
VERSION AF132478.1 GI:4378884
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3723)
Sengar,A.S., Wang,W., Bishop,V., Cohen,S. and Egan,S.B.
The EH and SH3 domain Eee proteins regulate endocytosis by linking
to dynamin and Eps15

Db 1441 TTGAAGCTCTGTAATGACAAAAAGCATGAGCTAAGAAAGAACTTCAGAGTATACAGGTGT 1500
Qy 1501 CCACTGGCAACCCAGAGGCAAGAAATTGAGAGGACGAAACAGTCTAGAGGCTAAGATTT 1560
Db 1501 CCACTGGCAACCCAGAGGCAAGAAATTGAGAGGACGAAACAGTCTAGAGGCTAAGATTT 1560
Qy 1561 GCTGAATACCCACTTAAAGCAGCAGTTGACAGAACTTCAGCAAAATGCTTGGAAAGATT 1620
Db 1561 GCTGAATACCCACTTAAAGCAGCAGTTGACAGAACTTCAGCAAAATGCTTGGAAAGATT 1620
Qy 1621 ATTCCAGAGAAACAGTACTCTGAGTACAGTTTAAACAAAGTCCAGAGAAACAGTTTGCAT 1680
Db 1621 ATTCCAGAGAAACAGTACTCTGAGTACAGTTTAAACAAAGTCCAGAGAAACAGTTTGCAT 1680
Qy 1681 AAGAGCTCGCTTCTTACCTCTCAAAAGAGCCTTGAAGCAAGAGAGCTGGCCGGCAGCAG 1740
Db 1681 AAGAGCTCGCTTCTTACCTCTCAAAAGAGCCTTGAAGCAAGAGAGCTGGCCGGCAGCAG 1740
Qy 1741 CTCGGGAGCAGCTGAGCAGAGGTGAGAGAGAGAACCGGTCAAAAGCTGAGAGATTTGAT 1800
Db 1741 CTCGGGAGCAGCTGAGCAGAGGTGAGAGAGAGAACCGGTCAAAAGCTGAGAGATTTGAT 1800
Qy 1801 GTTTTCAACAACAGCTGAAAGAACTGAGAGAGATTCATAGCAACAGCACTCCAGAA 1860
Db 1801 GTTTTCAACAACAGCTGAAAGAACTGAGAGAGATTCATAGCAACAGCACTCCAGAA 1860
Qy 1861 CAGAGGTCCCTGGAGCAGCGGCACTGAAGCAGAAAGAGCAGAGAGAGAGCCTGGAG 1920
Db 1861 CAGAGGTCCCTGGAGCAGCGGCACTGAAGCAGAAAGAGCAGAGAGAGAGCCTGGAG 1920
Qy 1921 TTAGAGAAACAAAAGGAGACGCTCAGAGACGATTCAGAGAAAGGAGCAAGCAATGGCTG 1980
Db 1921 TTAGAGAAACAAAAGGAGACGCTCAGAGACGATTCAGAGAAAGGAGCAAGCAATGGCTG 1980
Qy 1981 GAGCATGTGACAGAGAGAGAGAGCAGCCGCCCGGAAACCCCAAGAGAGAGAGAGAGAG 2040
Db 1981 GAGCATGTGACAGAGAGAGAGAGCAGCCGCCCGGAAACCCCAAGAGAGAGAGAGAGAG 2040
Qy 2041 AAGAGGAGAGACAGTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 AAGAGGAGAGAGAGTGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2101 GACAAAGCAGAGTGGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2160
Db 2101 GACAAAGCAGAGTGGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2160
Qy 2161 CCCTGGTCTTACACAGAGAAAGGCCGCTTACATTTCTGCAAGAGAGTGTAAAGTG 2220
Db 2161 CCCTGGTCTTACACAGAGAAAGGCCGCTTACATTTCTGCAAGAGAGTGTAAAGTG 2220
Qy 2221 GTATATTTACGAGCGCTGTATACCCCTTTGAATCCAGAAAGTCAAGATGATACCATCCAG 2280
Db 2221 GTATATTTACGAGCGCTGTATACCCCTTTGAATCCAGAAAGTCAAGATGATACCATCCAG 2280
Qy 2281 CCAGAGATATATGATCATGTGTGATGAAGACAGAGCTGAGAGCAGAGATGGCTTGGAGGA 2340
Db 2281 CCAGAGATATATGATCATGTGTGATGAAGACAGAGCTGAGAGCAGAGATGGCTTGGAGGA 2340
Qy 2341 GAGCTGAAAGGAGAGAGCGGATGTGTTCCCTGCAAACTATGAGAGAAAGATTCAGAAAT 2400
Db 2341 GAGCTGAAAGGAGAGAGCGGATGTGTTCCCTGCAAACTATGAGAGAAAGATTCAGAAAT 2400
Qy 2401 GAGGTTCACATCCAGCCAAACAGTGAACCGATCTGAATCTGCCCCCTCCCAAACTG 2460
Db 2401 GAGGTTCACATCCAGCCAAACAGTGAACCGATCTGAATCTGCCCCCTCCCAAACTG 2460
Qy 2461 GCTGTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTCTTGAAGCCCTCCAGAACCC 2520
Db 2461 GCTGTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTCTTGAAGCCCTCCAGAACCC 2520
Qy 2521 AACCACTGGGACAGCTTCAAGTTCACAGTGGCCAGAGCTCAAAAGAGAGAGAGAGAG 2580
Db 2521 AACCACTGGGACAGCTTCAAGTTCACAGTGGCCAGAGCTCAAAAGAGAGAGAGAGAGAG 2580

Qy 2581 GACCACTGGGATACGTGGGCGGCTCAGCCTTCTGACCGTACTAGTGTGCTGGCAGTTA 2640
Db 2581 GACCACTGGGATACGTGGGCGGCTCAGCCTTCTGACCGTACTAGTGTGCTGGCAGTTA 2640
Qy 2641 CGGCAAGATACGCTTTTAAACCCAGCAGCAGCAGTGGCTCTCCCATCTCCGCTCTG 2700
Db 2641 CGGCAAGATACGCTTTTAAACCCAGCAGCAGCAGTGGCTCTCCCATCTCCGCTCTG 2700
Qy 2701 GGGCAGAGTGAAGAGTGAAGGCTTACAGGCGCAGCCCTGTATTCCTTGGAGAGCCAA 2760
Db 2701 GGGCAGAGTGAAGAGTGAAGGCTTACAGGCGCAGCCCTGTATTCCTTGGAGAGCCAA 2760
Qy 2761 AAAAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAACAGCTTGGAGACAGAGAG 2820
Db 2761 AAAAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAACAGCTTGGAGACAGAGAG 2820
Qy 2821 ATGTGGTGTGTAAGAAAGTTCAGAGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAA 2880
Db 2821 ATGTGGTGTGTAAGAAAGTTCAGAGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAA 2880
Qy 2881 CTCATTTTCAAGGCGCGTAAAGAAATCCACAGCATGCTAGTGGCTTCTGAAAGTCT 2940
Db 2881 CTCATTTTCAAGGCGCGTAAAGAAATCCACAGCATGCTAGTGGCTTCTGAAAGTCT 2940
Qy 2941 GCTAGTCTTAAAGAGAGTGGCTTCCCGGCGGCAAGCCATTCGCGGAGAGAGATT 3000
Db 2941 GCTAGTCTTAAAGAGAGTGGCTTCCCGGCGGCAAGCCATTCGCGGAGAGAGATT 3000
Qy 3001 ATTGCAATGTACATACAGAGATTCAGAGCAAGAGATTTAACTTTTACAGCAAGGAGAT 3060
Db 3001 ATTGCAATGTACATACAGAGATTCAGAGCAAGAGATTTAACTTTTACAGCAAGGAGAT 3060
Qy 3061 GTGATTTGTGTTTACAAAGAAAGTGTGATCTGTGAGCGGAAACGCTGGGCAAGTCT 3120
Db 3061 GTGATTTGTGTTTACAAAGAAAGTGTGATCTGTGAGCGGAAACGCTGGGCAAGTCT 3120
Qy 3121 GGAAGTCTTCCCTTCTTACATGTAAGTGAAGGCTTAAAGTTTCAAGAGGCTTGGAACT 3180
Db 3121 GGAAGTCTTCCCTTCTTACATGTAAGTGAAGGCTTAAAGTTTCAAGAGGCTTGGAACT 3180
Qy 3181 AAAAGCAGGAGTTTGAAGAAACCTGAATTTGAGCCAGTATTTCTTCTACAGCCTGCT 3240
Db 3181 AAAAGCAGGAGTTTGAAGAAACCTGAATTTGAGCCAGTATTTCTTCTACAGCCTGCT 3240
Qy 3241 ACTGTGCCGAACAACCTGAGCTCTGAGGAGCTGATTCGATCCGGAAGAAAGAAC 3300
Db 3241 ACTGTGCCGAACAACCTGAGCTCTGAGGAGCTGATTCGATCCGGAAGAAAGAAC 3300
Qy 3301 CGAGGTGATGTGGGAGAGAGAACTGCAAGCTCGAGGAGAAAGCGCAGATAGGTTG 3360
Db 3301 CGAGGTGATGTGGGAGAGAGAACTGCAAGCTCGAGGAGAAAGCGCAGATAGGTTG 3360
Qy 3361 TTTCCAGCAAAATTAATGTAACCTTTAAGCCCGGAAACAGCAAAATCAACCCCACTGAG 3420
Db 3361 TTTCCAGCAAAATTAATGTAACCTTTAAGCCCGGAAACAGCAAAATCAACCCCACTGAG 3420
Qy 3421 CTAACCAAGACCGCAGTGCAGCAGCAGTGTGCTCAGGTATCGGATGTACATTAAC 3480
Db 3421 CTAACCAAGACCGCAGTGCAGCAGCAGTGTGCTCAGGTATCGGATGTACATTAAC 3480
Qy 3481 GGCAGAAACGATGAAGAACTAGCTTCAAGCAAAAGGCGCAGATATCAAGTCTCAACAG 3540
Db 3481 GGCAGAAACGATGAAGAACTAGCTTCAAGCAAAAGGCGCAGATATCAAGTCTCAACAG 3540
Qy 3541 GAGGACCCGAGCTGTGTAAGAGAGAGTCAAGTGGGCAAGTTGGCTCTTCCCATTCAT 3600
Db 3541 GAGGACCCGAGCTGTGTAAGAGAGAGTCAAGTGGGCAAGTTGGCTCTTCCCATTCAT 3600
Qy 3601 TATGTAAAGCTGACCAAGACATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 3642
Db 3601 TATGTAAAGCTGACCAAGACATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 3642

RESULT 2
AF132481 5145 bp mRNA linear ROD 09-MAR-1999
LOCUS Mus musculus Ebell protein mRNA, complete cds.
DEFINITION AF132481
ACCESSION AF132481
VERSION AF132481.1 GI:4378890
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Sengar, A.S., Wang, W., Cohen, S., Bishop, J., Cohen, S. and Egan, S.E.
The EH and SH3 domain Ebe proteins regulate endocytosis by linking
to dynamin and Eps15
JOURNAL EMBO J. 18 (5), 1159-1171 (1999)
MEDLINE 99164083
PUBMED 10064583
REFERENCE 2 (bases 1 to 5145)
AUTHORS Sengar, A.S., Wang, W., Cohen, S., Bishop, J. and Egan, S.E.
TITLE Direct Submission
JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood
Research/Developmental Biology, The Hospital for Sick Children, 555
University Avenue, Toronto, ON M5G-1X8, Canada
FEATURES
source
1..5145
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
1..5145
/codon_start=1
/product="Ebell protein"
/protein_id="AAB19749.1"
/db_xref="GI:4378891"
/translation="MAQPTPEFGSLDVAITVERAAXHDOFLSLKEIAGFITDQA
RNFPSGLPVLAIQIWLADMMNDGMDVEFIAKLKIKLQGLPSTLPVM
KOOPAISSAPFGIGIASMPPLTAVAPVMSGIPVAGMSPLVSSAPPAVPLAV
CAPVIOPLPAFPAHATWPPSSSTSPSSQANTKQXQKQSTVPAAPAAEMAYP
OSSLRKQLRNSHDKTSGHITGPQARTILMOSLPAQALASINLSDIDQGLTA
BEFLIAMLIDVANGQPLPVLPEYIIPSFRRVRSQMSVSISSSDVDLREPS
SEDEQPEKPLVTFEDKRENFERKQELRQALQOORKEOERLADQERAOER
KEREROEOAKROELEKOLEKORLEERREERKEIEREAAKRELERORLEMER
NRROELNORKEOEGTVLAKARKTLEFELEALNDKKHOLEKODIRGLATOROE
IESTKSEKLEIAETTHQQOQLESQGLSLPEKOLISLQKQVONSILHRSILT
LKRLAEKELAQLEBOLDEVERETKRLGEIDVFNNQLKELRHSKQQLQKQSL
EAARKQEOERKSLLEKOKEDQORVOERDKQMLEHVOOEORPRKPEERLKR
EDSVKKEAEERAKEMQDKSLRPHQEPKATLQAPMTEKGTPLTIAQESVY
VYRALYPERESKDEITIOGDIVVDSOTGEGMGLGELKGTGFWPANYAEKIP
ENEVPTPAKPVDTLSAPAKLALRETPAPLPVTSSESTPNMMADSSSPSSS
KPEIDNMWTMAQPSLTVPSAQQLRQSAFTPATRTGSSPVLQGEKREGLOAL
YPMRAKDNHNLNFKSDIVTALQEDMMFEBVOQKQWFKPSYKLSIGVRSSTSI
DTGPTESPASLKRVASPAKAPAIPEBEFIAMTYESSRQGLTFQOGDVIIVTKDGD
WMTGVGDKSGKGFPSNYRLKDSGEGTAGTGSIGKKEIAQVLSAAAGPEQTLT
APGOLILIRKKNPGMGEGLQARKKQIIGFPANYVYKLSIPGSKITTPLEKTA
QPAVCOVGMVDYTDONDELAFSGQIINLINKEDPMWKGVSQVGLPSNYVYK
TTNDPSSQWCSDDLHLDMLTTERKRGYIHELIVTEYNYNDQLTETIFQKPLTE
SELTREKAVAMIIVNMKELIMCNILKILALRRKMSSEKPMVKIIGILISAQLPHQ
PYIRFCSQQLNGALIQKTDPAEPKFEVKRLAMPCKGKPLSSFLIKMORTRY
PLIKNILENTENPNDHSLKHALEKELCSQVNEGVREKENSDBLEMTQAOVCE
GLSEOLVNSVTNCGPKRFLSGKLYAKSKNELYGLFENDELILLTITPLSGSGT
DKVSPKSNLQIKWYKTPILFNEVLYKLPDPSGDEPIFHSIDRIVYTLAEASINER
TAWOKITQASGLYIETEKREKAYIVRSQBATIGTGLMNVVGEILAKCRSHGKS
NPYCEVINGSOCHITKTIQDTLNPWNSNCOFFIRDLDEQVLCVIFERFDPSFDFL
GRTEIRVADIKDDQSGKPVTKCLHLHVPGEIIVRLDQLFDEP"

BASE COUNT 1448 a 1298 c 1415 g 984 t
ORIGIN

Query Match 99.9%; Score 3639.4; DB 10; Length 5145;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATGTCTGGGCGCATACCTGTGGAG 60

Db 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATGTCTGGGCGCATACCTGTGGAG 60
QY 61 GAAAGGCGCAGCATGACCCAGCACTTCTTAAGCTTGAAGCCGATAGCGGATTTATTA 120
Db 61 GAAAGGCGCAGCATGACCCAGCACTTCTTAAGCTTGAAGCCGATAGCGGATTTATTA 120
QY 121 GGTATCAGGAGAGAACTTTTTCATCTGGGTACCTCGACCTGCTTGAACA 180
Db 121 GGTATCAGGAGAGAACTTTTTCATCTGGGTACCTCGACCTGCTTGAACA 180
QY 181 ATATGGCGCTAGCGGACATGAATACATGAGATGAGATGATCAATGGAATTTTCATA 240
Db 181 ATATGGCGCTAGCGGACATGAATACATGAGATGAGATGATCAATGGAATTTTCATA 240
QY 241 GCCATGAAGCTTATCAACTGAAGCTTCAAGATATCACTCCCTCCACACTTCCCT 300
Db 241 GCCATGAAGCTTATCAACTGAAGCTTCAAGATATCACTCCCTCCACACTTCCCT 300
QY 301 GTCATGAACAGCAACGATGGCTATTTCCAGTGCACAGCATTTGGTATGAGGGATT 360
Db 301 GTCATGAACAGCAACGATGGCTATTTCCAGTGCACAGCATTTGGTATGAGGGATT 360
QY 361 GCTAGCATGCCACCACTCACAGCTGTGCTCTGTGCAATGGAGCTCCATTCCAGTTGT 420
Db 361 GCTAGCATGCCACCACTCACAGCTGTGCTCTGTGCAATGGAGCTCCATTCCAGTTGT 420
QY 421 GGAATGCTCCACCTTAGTATCTTCTGTCCCTCAGACAGATGCTCCCTGGCTAAC 480
Db 421 GGAATGCTCCACCTTAGTATCTTCTGTCCCTCAGACAGATGCTCCCTGGCTAAC 480
QY 481 GGGGCTCCTCCGCTATACAGCCCTGCTGCGCTTGGGATCTTGAGCCACATGGCA 540
Db 481 GGGGCTCCTCCGCTATACAGCCCTGCTGCGCTTGGGATCTTGAGCCACATGGCA 540
QY 541 AAGAGTTCTTCTTCAAGCAGATCTGTGCAAGGTCACAAATTAACATAAGTTTACAGAG 600
Db 541 AAGAGTTCTTCTTCAAGCAGATCTGTGCAAGGTCACAAATTAACATAAGTTTACAGAG 600
QY 601 GCAATATCATTCATGTGTCGCGAGCGCCCTTCACAGACAGAAATGGCTGTGCTCACTCA 660
Db 601 GCAATATCATTCATGTGTCGCGAGCGCCCTTCACAGACAGAAATGGCTGTGCTCACTCA 660
QY 661 TCAAGGCTGAATACAGGAGTTTATCAACAGCCAGCAAAATATGAGTGAAGCACTTA 720
Db 661 TCAAGGCTGAATACAGGAGTTTATCAACAGCCAGCAAAATATGAGTGAAGCACTTA 720
QY 721 ACAGGTCGCCAGGAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCACTGGCT 780
Db 721 ACAGGTCGCCAGGAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCACTGGCT 780
QY 781 TCAATATGGAATCTTTTGCATTCATGATGATGAGTGAAGAACTCACTGCAAGAAATTTATC 840
Db 781 TCAATATGGAATCTTTTGCATTCATGATGATGAGTGAAGAACTCACTGCAAGAAATTTATC 840
QY 841 CTACGTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 841 CTACGTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 CCAAGATATACCTCCT 960
Db 901 CCAAGATATACCTCCT 960
QY 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGAGGCTGCTGAGAGAGGCTGCTGAG 1020
Db 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGAGGCTGCTGAGAGAGGCTGCTGAG 1020
QY 1021 CCAAG 1080
Db 1021 CCAAG 1080
QY 1081 AGTGTGAGCTGAG 1140
Db 1081 AGTGTGAGCTGAG 1140

Db 1081 AGTGTGAGCTGAGAAAGCGCCGCAAGCGCTCTTGTGAGCAGCAGCGCAAAAGACAGAG 1140
Qy 1141 CGGTTGAGCTCAGCTGAGAGCGCGCGAGAGAGAGAAAGGAGGAGCGCCAGAGAG 1200
Db 1141 CGGTTGAGCTCAGCTGAGAGCGCGCGAGAGAGAGAAAGGAGGAGCGCGCCAGAGAG 1200
Qy 1201 GAGGCCAAGCGGAGCTGAGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAG 1260
Db 1201 GAGGCCAAGCGGAGCTGAGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAG 1260
Qy 1261 CAGCGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCGAGCGCCGCAAAACGAGAACTG 1320
Db 1261 CAGCGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCGAGCGCCGCAAAACGAGAACTG 1320
Qy 1321 GAAAGCAGAGCAACTGTAATGGGAAACGGAACCGGAGACGGAACCTCTGGAATTCAGAG 1380
Db 1321 GAAAGCAGAGCAACTGTAATGGGAAACGGAACCGGAGACGGAACCTCTGGAATTCAGAG 1380
Qy 1381 AACCAAGAGCAG 1440
Db 1381 AACCAAGAGCAG 1440
Qy 1441 TTGAAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1441 TTGAAAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Qy 1501 CGAGCTGGCAACCGAG 1560
Db 1501 CGAGCTGGCAACCGAG 1560
Qy 1561 GGTGAATATCAGCTTACAG 1620
Db 1561 GGTGAATATCAGCTTACAG 1620
Qy 1621 ATTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGTTAAACAGTTAAACAGTTAAAC 1680
Db 1621 ATTCCAGAGAAACAGATCTCAGTGAACAGTTAAACAGTTAAACAGTTAAACAGTTAAAC 1680
Qy 1681 AGAGATCGCTCTTACAG 1740
Db 1681 AGAGATCGCTCTTACAG 1740
Qy 1741 CTCGAG 1800
Db 1741 CTCGAG 1800
Qy 1801 GTTTTCAACCAACAGCTGAG 1860
Db 1801 GTTTTCAACCAACAGCTGAG 1860
Qy 1861 CAGAGGTCTCTGAG 1920
Db 1861 CAGAGGTCTCTGAG 1920
Qy 1921 TTGAGAAACAAAG 1980
Db 1921 TTGAGAAACAAAG 1980
Qy 1981 GAGCATGTGAG 2040
Db 1981 GAGCATGTGAG 2040
Qy 2041 AAGAGGAAAG 2100
Db 2041 AAGAGGAAAG 2100
Qy 2101 GACAAAG 2160
Db 2101 GACAAAG 2160
Qy 2161 CCTGTGTTACAG 2220
Db 2161 CCTGTGTTACAG 2220

Qy 2221 GTATATTACCGAGCGCTGTAACCCCTTTGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 GTATATTACCGAGCGCTGTAACCCCTTTGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy 2281 CCAGAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 2281 CCAGAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Qy 2341 GAGCTGAAG 2400
Db 2341 GAGCTGAAG 2400
Qy 2401 GAGGTTCCCACTCCAGCCCAACAG 2460
Db 2401 GAGGTTCCCACTCCAGCCCAACAG 2460
Qy 2461 GCTCTGCTGAG 2520
Db 2461 GCTCTGCTGAG 2520
Qy 2521 AACCACTGGGAG 2580
Db 2521 AACCACTGGGAG 2580
Qy 2581 GACAACTGGGAG 2640
Db 2581 GACAACTGGGAG 2640
Qy 2641 CGGAG 2700
Db 2641 CGGAG 2700
Qy 2701 GGCAG 2760
Db 2701 GGCAG 2760
Qy 2761 AAAAGCAACACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Db 2761 AAAAGCAACACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
Qy 2821 ATGT 2880
Db 2821 ATGT 2880
Qy 2881 CTCATTTCAG 2940
Db 2881 CTCATTTCAG 2940
Qy 2941 GCTAGCTTAAAG 3000
Db 2941 GCTAGCTTAAAG 3000
Qy 3001 ATTGCATGTACATACAG 3060
Db 3001 ATTGCATGTACATACAG 3060
Qy 3061 GTGATTTGTGTACCAAG 3120
Db 3061 GTGATTTGTGTACCAAG 3120
Qy 3121 GAGAGTCTTCCCTTCACTATGAG 3180
Db 3121 GAGAGTCTTCCCTTCACTATGAG 3180
Qy 3181 AAAACAGGAG 3240
Db 3181 AAAACAGGAG 3240
Qy 3241 ACTGTCTCCGAGCAACTCACTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3300
Db 3241 ACTGTCTCCGAGCAACTCACTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3300

QY 3301 CCAGGTGATGTTGGGAGAGAACTGCAAGCTGAGGGGAAAAAGCCGACAGTAGGGTGG 3360
 DB 3301 CCAGGTGATGTTGGGAGAGAACTGCAAGCTGAGGGGAAAAAGCCGACAGTAGGGTGG 3360
 QY 3361 TTTCGAGCAAAATTATGTCMAAATTCTTAAGCCCGGAAACAAATATCAACCCCACTGAG 3420
 DB 3361 TTTCGAGCAAAATTATGTCMAAATTCTTAAGCCCGGAAACAAATATCAACCCCACTGAG 3420
 QY 3421 CTACCCAGAACCCGAGTGCAGCCAGCAAGTGTGCCAGTGTGATCGGATGTACGATTACACC 3480
 DB 3421 CTACCCAGAACCCGAGTGCAGCCAGCAAGTGTGCCAGTGTGATCGGATGTACGATTACACC 3480
 QY 3481 GCCCAGAACGATGACGAACTGACCTTCAAGAAAGCCAGATCATCAAGCTCTCCAGCAAG 3540
 DB 3481 GCCCAGAACGATGACGAACTGACCTTCAAGAAAGCCAGATCATCAAGCTCTCCAGCAAG 3540
 QY 3541 GAGGACCCGAGCTGTGGAAAAGAGAACTGAGTGGGCAAGTTGGCTTCTCCATCCAT 3600
 DB 3541 GAGGACCCGAGCTGTGGAAAAGAGAACTGAGTGGGCAAGTTGGCTTCTCCATCCAT 3600
 QY 3601 TATGTAAAGCTGACACAGACATGAGACCCGACGACGATG 3641
 DB 3601 TATGTAAAGCTGACACAGACATGAGACCCGACGACGATG 3641
 RESULT 3
 AF127798 4025 bp mRNA linear ROD 28-JUN-1999
 LOCUS Rattus norvegicus EH- and SH3-domain containing protein EHSN1 mRNA,
 DEFINITION complete cds.
 ACCESSION AF127798
 VERSION AF127798.1 GI:4835852
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 4025)
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T.C.
 TITLE EHSN1, intersectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?
 JOURNAL J. Biol. Chem. 274 (26), 18446-18454 (1999)
 MEDLINE 99303509
 PUBMED 10373452
 REFERENCE 2 (bases 1 to 4025)
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T.C.
 TITLE Direct Submission
 JOURNAL Submitted (11-FEB-1999) Center for Basic Neuroscience and HMMI, Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
 FEATURES
 source
 location/Qualifiers
 1..4025
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 41..3694
 /codon_start=1
 /product="EH- and SH3-domain containing protein EHSN1"
 /protein_id="AAD30271.1"
 /db_xref="GI:4835852"
 /translation="MAQPPPTFGSLDITMAITVEBRAKHDOQFSLKPISEFTIGDQA
 RNFPGSLPQVLAQIWLADNKKDMDVFSIAKLKTLQGLPPLPVM
 KQDPAISAPAFIGGMAQMPPLTAVPVMSIPIVGMSPVSVPAVPLPV
 GAPVIOPLPAPAPATLPEKSSFSFGSGLNTKQKQSFVVASPAABWAV
 OSRLKROLFNSHDKTMSGHLTGFOATLIMOSLPOALASITWNSADIDOKLTA
 EEFILAMHLDIVASGOLPVLPEVYIPSPFRVRVSGMSVSSSSADRLPEES
 SEDQOYBKLPVTFEDKKNFRGNILEKRRQALBQKREKQERLAQLEKRAEGR
 KERERQERKQLELEKQERLEKQERLEKQERLEKQERLEKQERLEKQERLEKQER
 NRRELTORNKQDGIIVLKARKRTLEFELBANDKQHLQEGKLDRIKRLATQRE
 IESTNKSRELRIATLHLOQLOESQMLGILPEKQILSPOLQVONSLHRSLSLT
 LKRALKEKLARQLRQLRQLDEVEKETSRLQDEIVFNQKLKELRHSKQQLQKRSI

BASE COUNT 1147 a 1009 c 1123 g 746 t
 ORIGIN
 YVKLTIDMDSDSQ"
 EASRLKQERKSLLELEKQERGORVORBDKQOMOEHOVEQORPRKPHEDKLR
 EDSYKKEAEERAKREPVODKOSLFIHNGPAPDAQPMPTTEGGLTIAQESAKYV
 YKRLPFESRSHDEITQPDIVMAGVNDVSQTEPMLGSEPPKMGDPFANPA
 EKIPENEIPTPAPRPVLDLTAPAPKALRETPVLPVTSSSEPTTPNNMADFSWTS
 STNEKPEKTDMDWMAQPSLTVPSAQLRORSAPLTVATGSSSPSLVGGSEKVGQ
 AQAALYEMRKAQDNLNENKSDVITVLEQDMMFGEVQKQKPKSVYKLSIGPVK
 STSIDPTEPASPILKVASPAKPAIRGESEFAMLYESESHEDDLFFOQGHVWVK
 KQDQMWGTGSEGVSPNRYVRLKSESGCTAGKQSLKREKPEIAVTSYATGPE
 QLTAPQQLILIRKQNPGRWBEGLQARKRQIQGMFPANVYKLSIGTSKITTELP
 KTVAPQAVCCVIGMIDYTAQNDDELAFSKQIIVLNSKEDPDMWKESQVGLFSPN
 YVKLTIDMDSDSQ"
 Query Match 88.7%; Score 3229.8; DB 10; Length 4025;
 Best Local Similarity 93.2%; Pred. No. 0;
 Matches 3407; Conservative 0; Mismatches 232; Indels 18; Gaps 2;
 QY 1 ATGGCTAGTTTCCACACCTTTGCGTGTAGCCGTGATGTCGGGCAATATGAGAG 60
 DB 41 ATGGCTAGTTTCCACACCTTTGCGTGTAGCCGTGATGTCGGGCAATATGAGAG 100
 QY 61 GAAAGGCCAAGCATGACAGCAAGTCTTAGCCTGAAGCCGATAGCGGATTTATTA 120
 DB 101 GAAAGGCCAAGCATGACAGCAAGTCTTAGCCTGAAGCCGATATCTGATTCATCA 160
 QY 121 GGTGATCAAGGAGAACTTTTTCATCTGGGTTTACCTCAGCCTGTCTTACGCAA 180
 DB 161 GGTGATCAAGGAGAACTTTTTCATCTGGGTTTACCTCAGCCTGTCTTACGCAA 220
 QY 181 ATATGGCGCTAGCGGACATGAATPACATGAGAGATGATCAAGTGAATTTTCCATA 240
 DB 221 ATATGGCGCTAGCGGACATGAATPACATGAGAGATGATCAAGTGAATTTTCCATA 280
 QY 241 GCCATGAAGCTTATCAAACTGAAGCTTCAAGATATGAGTCTCCCTCCACATTCCT 300
 DB 281 GCCATGAAGCTTATCAAACTGAAGCTTCAAGATATGAGTCTCCCTCCACATTCCT 340
 QY 301 GTCATGAACAGCAACCAAGTGTATTTCCAGTGCACACGATTTGGTATGAGGAT 360
 DB 341 GTCATGAACAGCAACCAAGTGTATTTCCAGTGCACACGATTTGGTATGAGGAT 400
 QY 361 GCTAGCAGCCACACCTCAGCTGTGTCTCTGTGCAATGAGGCTCATTCAGTTGT 420
 DB 401 GCTAGCAGCCACACCTCAGCTGTGTCTCTGTGCAATGAGGCTCATTCAGTTGT 460
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGTCTCTCCACAGCAGTGTCTCCCTGGCT 480
 DB 461 GGAATGTCTCCACCTTATGATCTTCTGTCTCTCCACAGCAGTGTCTCCCTGGCT 520
 QY 481 GGGGCTCTCCGTCATACAGCTCTGCTGCTGCTGGCATCTCGACCAATGGCCA 540
 DB 521 GGGGCTCTCCGTCATACAGCTCTGCTGCTGCTGGCATCTCGACCAATGGCCA 580
 QY 541 AAGAGTCTCTCCACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 DB 581 AAGAGTCTCTCCACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640
 QY 601 GCACATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
 DB 641 GCACATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 700
 QY 661 TCAAGGCTGAATACAGGAGTATTTCAACAGCCAGCAAACTATGATGATGAT 720
 DB 701 TCAAGGCTGAATACAGGAGTATTTCAACAGCCAGCAAACTATGATGATGAT 760
 QY 721 ACAGGCTCCAGGAGCAAGTATTTCAATGATGATGATGATGATGATGATGATGAT 780
 DB 761 ACAGGCTCCAGGAGCAAGTATTTCAATGATGATGATGATGATGATGATGATGAT 820
 QY 781 TCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 821 TCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 880

QY	841	TTAGCTATGACACTAAATTGATGTTGTCACATGTCGTGATGACGCCATGCGCTCGTGCAT	900
Db	881	CTAGGAGTACACTGATTTGATGTTCTCTATGTCGTGATGACCCATGCTCGCTGTCTGCTGCAT	940
QY	901	CCAGAAATACATCCCTCTTCTCTTCCAGAAAGATTGCGTCCGGCAGTGAGATGTCGCTCATYA	960
Db	941	CCAGAAATACATCCCTCTTCTCTTCTTGAAGAGATTGCGTCCGGCAGTGAGATGTCGCTCATYA	1000
QY	961	AGCTCTTCTTCTGAGGATCAGAGGGTGGCTGAGGAGCCGTGCTGAGAGATGAGACAGAG	1020
Db	1001	AGCTTTTGTCTTGACAGACCAAGGGGTGGCCCGAGGAGCCATGCTGAGAGATGAGACAGAG	1060
QY	1021	CCAGAGAAAGAACTGCTCTGTGACATTTGAGATTAAGACGGGAGAACTTGAGCGAGAGC	1080
Db	1061	GTGAGAAAGAAAGCTGCTGTGTGACATTTGAGATTAAGAACGGGAGAACTTGAGAGAGAGC	1120
QY	1081	AGTGTGAGCTGTGAGAAACCGCCGCCACAGCGCTCTTGAGCAGCAGCGCAAAAGACAGAG	1140
Db	1121	AACCTGAGCTGTGAGAAACCGCAGGAGCGCTCTCTGTGAGCAGCAGCGCAAGAGCAGAG	1180
QY	1141	CGGTTGCTCAGCTGAGACCGCGCCGAGCAGGAGAGAGAGCGGGAGACGGCCAGGAGAG	1200
Db	1181	CGGCTGGCTCAGCTGTGAGCGCGCCGAGCAGAGAGAGAAAGCGGAGCGCCAGGAGAG	1240
QY	1201	GAGGCAAGCGGCAAGCTGAGACCTGAGAAAGCAGCTGAGAAACAGCGGAGACTGAGACGG	1260
Db	1241	GAGCGCAAGAGACAGCTGAGACTGAGAGAGCAGCTGAAAGACAGCGGAGCTGAGAGAG	1300
QY	1261	CAGGAGAGGAGGAGAGAGAGAGATTCGAGAGAGCGCGAGCGCCGCAAAACGGGAACTG	1320
Db	1301	CAGAGAGAGGAGAGAGAGAGAGATTCAGAGAGCGGAGCGAGCAAAACGGGAGCTC	1360
QY	1321	GAAAGGCGACGACCACTTGAATGGGAAACGGAACCGGACAGAGAACTCTGATATCAGAGG	1380
Db	1361	GAAAGGCGACGACAGCTTGAATGGGAAACGGAACCGGAACTCTGATATCAGAGG	1420
QY	1381	AACCAAGAGCAGAGAGGGCACCGTGTCTTGAAGGCAGAGAGAGAGACTTGTGAGTTTGAG	1440
Db	1421	AACCAAGAGCAGAGAGGGCATGTGTCTTGAAGGCAGAGAGAGAGACTTGTGAGTTTGAA	1480
QY	1441	TTAAGAGCTCGAATGACAAAGAGCATCAGCTAGAGAGAGAAACTTCAGGATATCAGGGT	1500
Db	1481	TTAAGAGCTCGAATGACAAAGAGCATCAGCTGAGAGAGAAACTTCAGGATATCAGGGT	1540
QY	1501	CGACTGGCAACCCAGAGGCAAGAAATTGAGACAGCAAGATCTAGAGACTAAGATT	1560
Db	1541	CGACTGGCAACCCAGAGGCAAGAAATTGAGACCAAAAGTCTAGAGACTGAGATT	1600
QY	1561	GCTGAATACACCCACTTACAGACAGATTGAGAGATTCAGCAAAATGCTTGGAAAGCTT	1620
Db	1601	GCGGAATACACCCACTTACAGCAACAGTTGAGAGATTCAGCAAAATGCTTGGAAAGCTT	1660
QY	1621	ATTCAGAGAAACAGATACAGTACGATGACAGTTAAACAAAGTCCAGACAGAAACAGTTTGAT	1680
Db	1661	ATTCAGAGAAACAGATATCTCAGTGACCAAGTTAAACAAAGTCCAGACAGAAACAGTTTGAT	1720
QY	1681	AGAGACTGCTTTTACTTACCTTCAAAAGAGCTTTGAAAGCAAGAGACTGCGCAGCAG	1740
Db	1721	AGAGATTCGCTTTTACTTACCTTCAAAAGAGCTTTGAAAGCAAGAGACTGCGCAGCAG	1780
QY	1741	CTCCGGGAGCAGCTGAGAGAGTGGAGAGAGACAGGCTCAAAAGCTGACAGAGATTTGAT	1800
Db	1781	CTTGAGAGAGCAGCTGAGAGAGTGGAGAGAGAGACAGGCTCAAAAGCTGACAGAGATTTGAT	1840
QY	1801	GTTTTCAACAAACAGCTGAAAGAACTGAGAGATACATAGCAAAACAGCACTCCAGAG	1860
Db	1841	GTTTTCAACAAACAGCTGAAAGAACTGAGAGATACATAGTAACAGCACTCCAGAG	1900
QY	1861	CAGAGGTCCTCGAGGAGCAGCGCGCATGAGAGCAAAAGCAGAGAGAGAGAAAGCTCTGAG	1920
Db	1901	CAGAGGTCATATCAGAGCGAGAGAGCTTAAGAGCAAAAGCAGAGAGAGAGAAAGCTCTGAG	1960

QY	1921	TTAAGAAAGCAAAAGAAAGCGCTCAGAGAGAGTTCAGGAAAGGACAAACAAATGCGTG	1986
Db	1951	TTGGAGAGCAAAAGGAAAGAGTTCAGAGCAAGTTCAGGAAAGGACAAACAAATGCGTG	20220
QY	1981	GAGCATGTGCAGCAGAGAGCAGCCAGCCCGGAAACCCCAAGAGAGACGACTG	20408
Db	2021	GAGCATGTGCAGCAGAGAGAGCAGCAGCGCTCGGAAACCCCAAGAGAGACAAATGTG	20808
QY	2041	AAGAGGGAAGCACTGTTCAGGAAAGAGAGCGGAAAGAGAGCCAGCCGGAATGCA	21000
Db	2081	AAAAGGGAAGCACTGTTCAGGAAAGAGAGCGGAAAGAGAGCCAGCCGGAATGCA	21408
QY	2101	GACAAGCAGAGTCGGCTTTTCCATCCCGCATCAGAGCAGCTAAGCTGGCCACCAGGCA	21668
Db	2141	GACAAGCAGAGTCGGCTTTTCCATCCTCATCAGAGCAGCTAAGCCG---CCGAGGCA	21978
QY	2161	CCCTGTGTCTACACAGAGAAAGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTG	22208
Db	2198	CCCTGTGTCTACACAGAGAAAGGTCGCTTACCATTTCTGCACAGAGAGTGTAAAGTG	22578
QY	2221	GTATATTAACGAGCGCTGTACCCCTTTGAATCCAGAACTCAGATGAGATCCATCCAG	22808
Db	2258	GTATATTAACGAGCGCTGTACCCCTTTGAAGTCCAGAACTCATGACGATCACATCCAG	23178
QY	2281	CCAGAGATATATGATC-----GGTGAATGAATACCCAGACTGAGAGACCA	23258
Db	2318	CCAGAGATATATGATCAGTTGTAAGGGGAATGGGTGATGAATACCCAGACCGAGAGACCA	23778
QY	2326	GGATGCGCTTGAGAGAGAGCTGAAAGGGAAGACGAGATGTGTCCTGCAAACTATGACGAA	23858
Db	2378	GGATGCGCTTGAGAGAGAGACCGAGGGGGAAGACGAGATGTGTCCTGCAAACTATGACGAG	24378
QY	2386	AAGATTCAGAAATGAGGTTCCCACTCCAGCCCAACCACTGACCGATCTGACACTGCC	24458
Db	2438	AAGATTCAGAAATGAGATTCGCCACTCCAGCCCAACCACTGACCGATCTGACACTGCC	24978
QY	2446	CTTCGCCCCCAACTGTGCTCTGCGGTGAACCCCTGCTCTTTGGCCAGTGAACCTCTTCGAG	25058
Db	2498	CTTCGCCCCCAACTGTGCTCTGCGGTGAACCCCTGCTCTTTGGCCAGTGAACCTCTTCGAG	25578
QY	2506	CCCTCCCAACCCCCCAACTGCGGAGACTTGAGTCCAGCGTGGCCAGAGCTCAAC	25658
Db	2558	CCCTCCCAACCCCCCAACTGCGGAGACTTGAGTTCAGCGTGGCCAGAGCAAAAT	26178
QY	2566	GAGAGCCAGAAACGGAACACTGGATACGTGGCGGCTCAGCCTTCTGACCCGTAAT	26258
Db	2618	GAGAAACAGAAACGGAACACTGGGACACGTGGCGGCTCAGCCTTCTGACCTGTAACC	26778
QY	2626	AGTGTGCGCAATTCGCGCAGAGATCAGCCTTTACCCAGCAGACGCACTGCTCTCC	26858
Db	2678	AGTGTGCGCAATTCGCGCAGAGAGGTCCGCTTCAACAACGCAACGCACTGCTCTCC	27378
QY	2686	CCATCTCCCGTCCGTGGGCGAGGGGTGAAGGTGGAAGGGCTACAAGGCAAGCCCTGTAT	27458
Db	2738	CCCTTCTCCCGTCTGTGGCGCAGGGGTGAAGGGCTCCAGACAGACCTGTAT	27978
QY	2746	CCCTGAGAGCCAAAAAGAACCACTTAATTTTAAACAAAGTGAAGTCAACCGTT	28058
Db	2798	CCCTGAGAGCCAAAAAGAACCACTTAATTTTAAACAAAGTGAATGTCATACCGTT	28578
QY	2806	CTGGAACAGCAGACATGTGTGTGTTGGAAGATTCAAGTCCAGAAAGGTTGTGCC	28658
Db	2858	CTGGAACAGCAGACATGTGTGTGTTGGAAGATTCAAGTCCAGAAAGGTTGTGCC	29178
QY	2866	AAGCTTAACGTGAATCTCATTTTCAAGGGCCCGTAAAGAAATCCACACATTCGATCTGGC	29258
Db	2918	AAGCTTAACGTGAATCTCATTTTCAAGGGCCCGTAAAGAAATCCACACATTCGATCTGGC	29778
QY	2926	CTTACTAAAGCTCGTGTAGTCTTAAAGAGTGGCTTCCCGGCGGCAAGCCAGCATT	29858
Db	2978	CTTACTAAAGCTCGTGTAGTCTTAAAGAGTGGCTTCCCGGCGGCAAGCCAGCATT	30378
QY	2986	CCCGAGAGAGATTATTTGCGCATGTACACATTCAGAGATTCTGAGCAGAGATTAAAC	30458

Dd		3038	CCTGGAGAAAGAATTGTGTCCTATGTGACCTTACGAGAGTTCTGACCAATGAGATTTTAACC	3097
Oy		3046	TTTCAGCAAGGGAGTGATATTGTGTGTTACAAGAAAAGTGTGACTGTGTGACCGGAA CG	3105
Dd		3098	TTTCAGCAAGGGCATGTGATGTGTGTGTTACAAGAAAGATGTGGA CTGGTGGACGGGCACA	3157
Oy		3106	GTTGGCCGACAAAGTCCGAGAGTCTTCCCCTTCTAATACTGTGAGGCTTAAAGATT CAGAGGCC	3165
Dd		3158	GTTGGCCGAGAGCTCCGAGAGTCTTCCCCTTCTAATACTGTGAGGCTTAAAGACTCAGAGGCC	3217
Oy		3166	TCTGAACTGCTGGGGAAAAACAGGAGTTTAGGAAAAAACCTGAAATTGCCAGGTTATT	3225
Dd		3218	TCTGAACTGCTGGGGAAAAACAGGAGTTTAGGAAAAAACCTGAAATGCCAGGTTATT	3277
Oy		3226	GCTTCTCAAGCTGCTACTGTGTCCGAAACAATCACCTTGCTCTGGGACGCTGATTTCTG	3285
Dd		3278	GCTTCTCAAGCTGCTACTGTGTCCGAAACAATCACCTTGCTCTGGGACGCTGATTTTG	3337
Oy		3286	ATCCGAAAAAAGAACCCAGAGTGTGATGTGGGAGAGAACTGCAAAGCTTCAGGAAAAAG	3345
Dd		3338	ATCCGAAAAAAGAACCCAGAGTGTGATGTGGGAGAGAACTGCAAAGCTTCAGGAAAAAG	3397
Oy		3346	CGCCAGATAGGGTGTGTTCCAGCAAATTAATGCAAACTTCTAAGCCCGGAAACAAGCAA	3405
Dd		3398	CGCCAGATAGGGTGTGTTCCAGCAAATTAATGCAAACTTCTAAGCCCTGGAACAAGCAA	3457
Oy		3406	ATCACCCCAACTGAGCTACCCCAAGACCGAGTGCAGCAGCAGTGTGCAGGTGATCGGG	3465
Dd		3458	ATCACCCCAACTGAGCTACCCCAAGACCGAGTGCAGCAGCAGTGTGCAGGTGATCGGG	3517
Oy		3466	ATGTATGATTTACACCGCCAGAACATGACGATAGTACCTTACGAAAGCCAGATCATC	3525
Dd		3518	ATGTATGATTTACACCGCCAGAACATGACGATAGTACCTTACGAAAGCCAGATCATC	3577
Oy		3526	AACGTCCTCAACAAGAGGACCCCGAGTGTGTGAAAGAGAAAGTACATGTGGCAAGTTGGG	3585
Dd		3578	AACGTCCTCAACAAGAGGACCCCGAGTGTGTGAAAGAGAAAGTACATGTGGCAAGTTGGG	3637
Oy		3586	CTCTTCCCATCCAAATTATGTAAAGCTGACCAACAGACATGSAACCCAGCAGCAATGA	3642
Dd		3638	CTCTTCCCATCCAAATTATGTAAAGCTGACCAACAGACATGSAACCCAGCAGCAATGA	3694
RESULT 4				
LOCUS	AF114488	5381 bp	mRNA	linear PRI 16-JUL-2002
DEFINITION	Homo sapiens interseclin short isoform (ITSN)		mRNA,	complete cds.
ACCESSION	AF114488			
VERSION	AF114488.1	GI:4808924		
KEYWORDS	.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 5381) Pucharcos,C., Fuentes,J.J., Casas,C., de la Luna,S., Alcantara,S., Arbones,M.L., Soriano,E., Bistiroll,X. and Pritchard,M. Alu-splice cloning of human interseclin (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down syndrome Eur. J. Hum. Genet. 7 (6), 704-712 (1999)			
TITLE				
JOURNAL				
MEDLINE				
PUBMED	10482960			
REFERENCE	2 (bases 1 to 5381)			
AUTHORS	Pucharcos,C., Fuentes,J.J., Pritchard,M. and Bistiroll,X. Direct Submission Submitted (16-DEC-1998) Medical and Molecular Genetics Center, Cancer Research Institute, U'Hospitalet de l'lo., Avia. Castelldefels km. 2,7, Barcelona 08907, Spain			
JOURNAL				
TITLE				
FEATURES				
Source	1..5381			/organism="Homo sapiens"

[illegible]

QY 481 GGGGCTCTCCCTCATACAGCCTCTGCTGCTTGGCGATCTTCAGCCACATGGCCA 540
DB 749 GGGGCTCTCCCTGTATACAACTCTGCTGATTTGCTATCTTCAGCCACATGGCCA 808
QY 541 AAGAGTTCTTCTTCAGCAGATTTGGTCCAGGGTCAATTAACCTTAAGTTACAGAG 600
DB 809 AAGAGTTCTTCTTGTAGATTTGGTCCAGGGTCAATTAACCTTAAGTTACAGAG 868
QY 601 GCACATCATTTGATGTCAGAGGCGGCTCCAGCAGCAAGAGGGCTGGCTCATGTA 660
DB 869 GCACATCATTTGATGTCAGAGGCGGCTCCAGCAGCAAGAGGGCTGGCTCATGTA 928
QY 661 TCAGAGCTGAATACAGAGATTTATTCACAGCCACAGCAAACTATGATGACACTTA 720
DB 929 TCAGAGCTGAATACAGAGATTTATTCATATGTCAGCAAACTATGATGACACTTA 988
QY 721 ACAGGTTCCCGAGCAAGAACTATTTCAATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780
DB 989 ACAGGTTCCCGAGCAAGAACTATTTCAATGCAATGCAAGTTTACCCAGGCTCAGTGGCT 1048
QY 781 TCAATATGAAATCTTCTGACATTTGATCAAGATGAGAACTCACTGACAGAAATTTATC 840
DB 1049 TCAATATGAAATCTTCTGACATTTGATCAAGATGAGAACTTACAGCAGAGAAATTTATC 1108
QY 841 CTAGCTATGCACTTAATGATTTGTCATGTCGTCAAGCACTGCGCCGCTGCTGCT 900
DB 1109 CTGGAATGCACTTATGATTTGATGATGTCGTGCGCAACCTGCTGCTGCTGCT 1168
QY 901 CCAAAATACATCCCTCTCTCTTCTTCAAGAGATTCGCTCCGCGAGTGGATGTCCTCATTA 960
DB 1169 CCAAAATACATCCCTCTCTCTTCTTCAAGAGATTCGATCTGCGAGTGGATGTCCTCATTA 1228
QY 961 AGCTCTTCTTCTGTCATGAGAGGCTGCTGAGAGCCGCTGTCAGAGATGAGCAGAG 1020
DB 1229 AGCTCTTCTTCTGTCATGAGAGGCTGTCAGAGAGCCAGTTTCAAGATGAGCAAA 1288
QY 1021 C---CAGAGAGAACTGCTGTCATTTGAATTAAGAGCGGAGAACTTGCAGCA 1077
DB 1289 CAATTAGAAAAGAAATTAATCTGTTAAGTTGAAGTAAGAGGAGAACTTTGAACGT 1348
QY 1078 GGCATGTCAGAGTCGAGAGAGCGCGCCAGAGCGCTCTTGAAGCAGAGCGCAAGAGCAG 1137
DB 1349 GGCATGTCGAACTGAGAGAAAGCAAGAGCGCTCTGAGAACAGAGCGCAAGAGCAG 1408
QY 1138 GAGCGTTGCTCTGATGTCAGAGCGCGCCAGCAGAGAGAGAGAGAGAGAGAGAGAG 1197
DB 1409 GAGCGCTGCTGCTGATGTCAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAG 1468
QY 1198 CAGAGAGCGCAAGCGCAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAG 1257
DB 1469 CAGAGAGCGCAAG 1528
QY 1258 CCGCAGCAG 1317
DB 1529 CCGCAG 1588
QY 1318 CTGGAAGAGCAG 1377
DB 1589 CTGGAAGAGCAG 1648
QY 1378 AGGAAACAAG 1437
DB 1649 AGGAAACAAG 1708
QY 1438 GAGTTAGAGCTTCTGATGACAAAAGCATCAGCTAGAGAGAGAGAGAGAGAGAG 1497
DB 1709 GAGTTAGAGCTTCTGATGATTAAGAGCATCACTAGAGAGAGAGAGAGAGAGAGAT 1768
QY 1498 TGTGCACTGAGCAAG 1557
DB 1769 TGTGCACTGAGCAAG 1828
QY 1558 ATTGCTGAATCAACCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617

DB 1829 ATTGCGAAATCAACCATCTACAGCAAACTTACAGAGATCTCAGAAATGCTTGGAGAA 1888
QY 1618 CTTATTTCCAGAGAAACAGATTAATCTCAGTGCAGCTTAAACAGTCCAGAGAACTTTG 1677
DB 1889 CTTATTTCCAGAGAAACAGATTAATCTCAGTGCAGCTTAAACAGTTCAGAGAACTTTG 1948
QY 1678 CATAGAGATCGCTCTTACCTTCAAAAGAGCTTGAAGCAAGAGAGAGAGAGAGAGAG 1737
DB 1949 CATAGAGATTCAGCTTTTCACTTTTCACTTAAGAGCTTGAAGCAAGAGAGAGAGAG 2008
QY 1738 CAGCTCCGAGCAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1797
DB 2009 CAGCTTCAAGAT 2068
QY 1798 GATGTTTCAACACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
DB 2069 GATGTTTCAACACAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
QY 1858 AAGCAGAGTCCCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
DB 2129 AAGCAGAGTCCCTGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2188
QY 1918 GAGTTAGAT 1977
DB 2189 GAGTTAGAT 2248
QY 1978 CTGAGCATGTCAGAGAGAG---GAGCAGCAGCGCCCGGAGAAAGAGAGAGAGAG 2034
DB 2249 CTGAGCATGTCAG 2308
QY 2035 AGACTGAAG 2094
DB 2309 AGACTGAAG 2368
QY 2095 ATGCAAGACAG 2154
DB 2369 ATGCAAGACAG 2428
QY 2155 CAGGCACTGCTGTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
DB 2429 CAGGCACTGCTGTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2488
QY 2215 AAGAGTGAATTAATCAG 2274
DB 2489 AAGAGTGAATTAATCAG 2548
QY 2275 ATCCAGCCAGAGATTAATGATCAT-----GGTGAATGAAGAGAGAGAGAGAG 2319
DB 2549 ATCCAGCCAGAGAGATTAATGATCAT-----GGTGAATGAAGAGAGAGAGAGAG 2608
QY 2320 GAGCAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2379
DB 2609 GAGCAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2668
QY 2380 GAGAGAGAGATTCAG 2439
DB 2669 GAGAGAGAGATTCAG 2728
QY 2440 TCTGCCCCCTGCCCCCAAACTGGCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAG 2499
DB 2729 TCTGCCCCCTGCCCCCAAACTGGCTCTGCGTGAAGAGAGAGAGAGAGAGAGAGAG 2788
QY 2500 TCTGAGCCCTCCAGAACCCCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2559
DB 2789 TCTGAGCCCTCCAGAACCCCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2848
QY 2560 TCAAG 2619
DB 2849 TCAAG 2908
QY 2620 GTACCTAGTGTGAG 2679

Best Local Similarity 86.5%; Pred. No. 0;
Matches 3166; Conservative 0; Mismatches 475; Indels 21; Gaps 3;

QY 1 ATGGCTCAGTTTCCCAACCTTTCGGTGTGATGCTGGGCTCATTAAGTGGAG 60
DB 269 ATGGCTCAGTTTCCCAACCTTTCGGTGTGATGCTGGGCTCATTAAGTGGAG 328
QY 61 GAAAGGGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 329 GAAAGGGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 388
QY 121 GGTGATCAAGGAGGAACTTTTTCCTGAGTTCAGTGGCTGCTGCTTACGACA 180
DB 389 GGTGATCAAGTGGAACTTTTTCCTGAGTTCAGTGGCTGCTGCTTACGACA 448
QY 181 ATATGGGCGCTAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 449 ATATGGGCGCTAGCGGATGATGATGATGATGATGATGATGATGATGATGAT 508
QY 241 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATTCAGTCCCTCCACACTTCC 300
DB 509 GCTATGAATCTTATCAAACTGAAGCTACAGAGATTCAGTCCCTCCACACTTCC 568
QY 301 GTCTATGAACAGCAACCAAGTGTATTCAGTGCACAGCAATTTGATAGAGGAT 360
DB 569 GTCTATGAACAGCAACCAAGTGTATTCAGTGCACAGCAATTTGATAGAGGAT 628
QY 361 GCTATGAATGCAACCAAGTGTATTCAGTGCACAGCAATTTGATAGAGGAT 420
DB 629 GCTATGAATGCAACCAAGTGTATTCAGTGCACAGCAATTTGATAGAGGAT 688
QY 421 GGAATGCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGGCTCCCTGAGTAA 480
DB 689 GGAATGCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGGCTCCCTGAGTAA 748
QY 481 GGGGCTCTCCCTGATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 749 GGGGCTCTCCCTGATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 808
QY 541 AAGAGTTCTTCTTCAAGCAGATCTGCTCCAGGCTCACAATTAACCTAATTA 600
DB 809 AAGAGTTCTTCTTCAAGCAGATCTGCTCCAGGCTCACAATTAACCTAATTA 868
QY 601 GCACATCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
DB 869 GCACATCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 928
QY 661 TCAAGGCTGAATACAGGAGTATTCACAGCCAGCAAACTATGATGAGCACTTA 720
DB 929 TCAAGGCTGAATACAGGAGTATTCACAGCCAGCAAACTATGATGAGCACTTA 988
QY 721 ACAGGCTCCCAAGCAAACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGG 780
DB 989 ACAGGCTCCCAAGCAAACTATTCATGCAATCAAGTTTACCCAGGCTCAGTGG 1048
QY 781 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 1049 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1108
QY 841 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATG 900
DB 1109 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATG 1168
QY 901 CCAAGATACATCCCTCTCTTCAAGAGATTCGCTCGGAGTGGAGATGCTGCTAT 960
DB 1169 CCAAGATACATCCCTCTCTTCAAGAGATTCGCTCGGAGTGGAGATGCTGCTAT 1228
QY 961 AGCTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
DB 1229 AGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1288
QY 1021 C---CAGAGAGAACTGCTGTGATGATGATGATGATGATGATGATGATGAT 1077

DB 1289 CAATTGAAAAGAAATTTACCTGTAACTGTTGAAGATTAAGACGGGAGAACTTTGAACGT 1348
QY 1078 GGCAGTGTGACTGTGAAGAGCGCCGCAAGGCTCTTTGAGCAGAGAGGCAAGAGCG 1137
DB 1349 GGCAGTGTGACTGTGAAGAGCGCCGCAAGGCTCTTTGAGCAGAGAGGCAAGAGCG 1408
QY 1138 GAGCGGTTGCTCAGCTGTGAGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
DB 1409 GAGCGGTTGCTCAGCTGTGAGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1468
QY 1198 CAGAGGCCAAGCGGAGCTGTGAGAGAGCAGCTGTGAGAGAGCAGCGGAGCTGTGAG 1257
DB 1469 CAGAGGCCAAGCGGAGCTGTGAGAGAGCAGCTGTGAGAGAGCAGCGGAGCTGTGAG 1528
QY 1258 CCGCAGCGAG 1317
DB 1529 CCGCAGCGAG 1588
QY 1318 CTGAAAAGGCGAGCAACTTGAATGGGAAAGGAAACCGGAGCAGGAACTCTGAAATCAG 1377
DB 1589 CTGAAAAGGCGAGCAACTTGAATGGGAAAGGAAACCGGAGCAGGAACTCTGAAATCAG 1648
QY 1378 AGGAAACAGAGCAGAGAGGCAACGCTGTCTGAAAGCAGAGAGAGAGAGAGAGAG 1437
DB 1649 AGGAAACAGAGCAGAGAGGCAACGCTGTCTGAAAGCAGAGAGAGAGAGAGAGAG 1708
QY 1438 GAGTTTGAAGCTCTGAATGACAAAAGCTTCACTTGAAGAGAGAGAGAGAGAGAG 1497
DB 1709 GAGTTTGAAGCTCTGAATGACAAAAGCTTCACTTGAAGAGAGAGAGAGAGAGAG 1768
QY 1498 TGTGACCTGGCAACCAAGAGGCAAAATTTGAGAGACAGAACTGTGAGAGCTTAA 1557
DB 1769 TGTGACCTGGCAACCAAGAGGCAAAATTTGAGAGACAGAACTGTGAGAGCTTAA 1828
QY 1558 ATTGTGAATCAACCACTTACAGAGCAGCTTGAAGATCTCAGCAAAATGCTTGAAGA 1617
DB 1829 ATTGTGAATCAACCACTTACAGAGCAGCTTGAAGATCTCAGCAAAATGCTTGAAGA 1888
QY 1618 CTTATTTCCAGAGAAACAGATATCTCAGTGAACCACTTGAAGAGAGAGAGAGAG 1677
DB 1889 CTTATTTCCAGAGAAACAGATATCTCAGTGAACCACTTGAAGAGAGAGAGAGAG 1948
QY 1678 CATGAGATCGCTCTTTCACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCGAG 1737
DB 1949 CATGAGATCGCTCTTTCACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCGAG 2008
QY 1738 CAGCTCCGAGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
DB 2009 CAGCTCCGAGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2068
QY 1798 GATGTTTCAACCAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
DB 2069 GATGTTTCAACCAAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128
QY 1858 AAGCAGAGGCTCCGAGAGCAGGAGCTGGAAGCAGAAAGAGCAGAGAGAGAGAGAG 1917
DB 2129 AAGCAGAGGCTCCGAGAGCAGGAGCTGGAAGCAGAAAGAGCAGAGAGAGAGAGAG 2188
QY 1918 GAGTTTGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
DB 2189 GAGTTTGAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2248
QY 1978 CTGAGCATGTGACAGAGAG--GAGCAGCAGCGCCCGGAAACCCAGAGAGAGAG 2034
DB 2249 CTGAGCATGTGACAG 2308
QY 2035 AAGCTGAAG 2094
DB 2309 AAGCTGAAG 2368
QY 2095 ATGCAAGACAG 2154
DB 2369 GCACAAAGACAG 2428

QY	2155	AAGGCAACCTTGATCTACCAACAGAGAAAGACCCCGCTTACATTTCTGCAACAGAGAGTGA	221.4
Db	2429	CAGGCAACCTTGATCTACCAACAGAGAAAGAGTCACTTACATTTCTGCAACAGAGAAATGA	248.8
QY	2215	AAAGTGTATATTAACGAGCGCTGTACCCCTTTGAATCAGAAAGTCAGATGAGATCAC	227.4
Db	2489	AAAGTGTATATTAACCGGCACTGTACCCCTTTGAATCAGAAAGTCAGATGAGATCACT	254.8
QY	2275	ATCCAGCCAGAGAGATATAGTCAAT-----GGTGAATGAAGGCAGACTGGGA	231.9
Db	2549	ATCCAGCCAGAGAGATATAGTCAATGATTAAAGGGAGATGGGTGAATGAAGGCAGAACTGGA	260.8
QY	2320	GAGCCAGAGATGGCTTGAGAGAGCTGAAAGGAGAGCGGATGGTTCCTGCAAACTAT	237.9
Db	2609	GAAACCCGCTGGCTTGAGAGAGAAATTAAGAGAAAGACAGGATGGTTCCTGCAAACTAT	266.8
QY	2380	GCAGAAAAGATTCAGAAAATGAGGTTCCCACTCCAGCCAAACAGTAGACCGATCTGACA	243.9
Db	2669	GCAGAGAAAATTCAGAAAATGAGGTTCCCGCTCAGAGTAAACCGTAGTACATGATTCACACA	272.8
QY	2440	TCTGCCCCCTGGCCCCCAAACTGGGCTCTGGGTGAGACCCCTGCTCCCTTTCAGAGTACCTT	249.9
Db	2729	TCTGCCCCCTGGCCCCCAAACTGGGCTCTGGGTGAGACCCCGCTCCCTTTCAGATACCTT	278.8
QY	2500	TCTGAGCCCTTCACAAACCCCAACCAACTGGGCAGCTTCAAGTTCACGCTGACCCAGCAGC	255.9
Db	2789	TCAGAGCCCTTCACAGACCCCTTAATTAATGAGGCCGACCTTCACTCACTCACTGACCCACAGC	284.8
QY	2560	TCAAAACGAGAAACCGAGAACGAGCAACTGGAGTANAGTGGGGGGCTCAAGCCTTCTGTGAC	261.9
Db	2849	ACGAATGAGAAACCGAGAACGAGTAACTGGAGTGAATGGGACCGCCACCTCTCTCAAC	290.8
QY	2620	GTACCTAGTGTGGGCAAGTAAACGAGAGATCAGGCTTATACCCAGCCACAGCACTGGC	267.9
Db	2909	GTTCCAAGTGTGGGCAAGTAAAGGAGAGTGTCCGCTTATCTCAAGCCACGGCACTGGC	296.8
QY	2680	TCTCTCCCATCTCCCGTCTTGGGCCAGGGTGAAAGGTGAAAGGCTTACAGCCGACGC	273.9
Db	2969	TCTCTCCCGTCTCTCGTGTCTAGGCGCAGGGGTGAAAGGTGAGGGGCTTACAGGCTCAAGCC	302.8
QY	2740	CTGTATCCCTGAGAGCCAAAAAGACAAACCACTTAATTTTAAACAAAAGTGAAGTCACTC	279.9
Db	3029	CTATATCTTGTGAGAGCCAAAAAGACAAACCACTTAATTTTAAACAAAATGATGTCACTC	308.8
QY	2800	ACCGTCTCGAAACAGCAAGACATGTGTGTGTGTGGAGAGTTCAAGGTCAGAGGGTTGG	285.9
Db	3089	ACCGTCTCGAAACAGCAAGACATGTGTGTGTGTGGAGAGTTCAAGGTCAGAGGGTTGG	314.8
QY	2860	TTCCCAAGTCTTANAGTAACTCATTTCAAGGCCCGCTTAAGAGAAATCCAGAACATGCAT	291.9
Db	3149	TTCCCAAGTCTTANAGTAACTCATTTCAAGGCCCGCTTAAGAGAAATCTACAGCATGGAT	320.8
QY	2920	ACTGGCCCTACTGAAAGTCTGTGATCTTAAAGAGAGTGCGTTCGCCGCGCCAGCAAGCA	297.9
Db	3209	TCTGGTTCCTTACAGAGAGTCTGTGATCTTAAAGAGAGTGAAGCTCTCCAGCAGCCAGACGC	326.8
QY	2980	GCCATTCCTCGAGAAAGATTATTTGCCATGTACATACACAGAGTCTTGACAGAGAGAT	303.9
Db	3269	GTCGTTTCGGAGAGAAATTAATTCCTCACTTACAGAGTCTTGACAGAGAGAT	332.8
QY	3040	TTAACCTTTACAGCAAGGGATGTGTATGTGGTTTACCAAGAAAGATGTGACTGTGGAGCG	309.9
Db	3329	TTAACCTTTACAGCAAGGGATGTGTATGTGGTTTACCAAGAAAGATGTGACTGTGGAGCA	338.8
QY	3100	GGAACGCTGGGCGACAAAGTCGAGACTTTCCTTTCTAACTATGTGAGGCTTAAAGATTCA	315.9
Db	3389	GGAACAGTGGGCGACAAAGTCGAGACTTTCCTTTCTAACTATGTGAGGCTTAAAGATTCA	344.8
QY	3160	GAGGCTCTTGAACCTGTGGGAAAAACAGGAGTTTAAAGAAAAAAACCTGAAGTTGCCAG	321.9
Db	3449	GAGGCTCTTGAACCTGTGGGAAAAACAGGAGTTTAAAGAAAAAAACCTGAAGTTGCCAG	350.8

QY	3220	GTAAATGGTTCCTACAGCTGACTGATCGATCCGAAACAATCAACCCCTGGCTCCGGGACACTG	3279
Db	3509	GTTATATGCTCTATACACCGCCACCGGCCCCGAGACGTCACCTCGCCCCCTGGTCACTG	3568
QY	3280	ATTCTGATCCGGAAAAAGAACCCAGGTGATGTGGAGAGAGACTGCAAAGCTTCGAGGG	3339
Db	3569	ATTATTGATCCCAAAAAAGAAACCCGAGTGATGTGGAGAGAGAGCTGCAAAGCACTGTGG	3628
QY	3340	AAAAAGGCCAGATAGGGTGGTTCCAGCAAAATTATGTCAAACTCTTAGCCCCGGAA	3399
Db	3629	AAAAAGGCCAGATAGGGTGGTTCCAGCTAATTATGTAAGCTTCTTAAGCCCTGGAGCG	3688
QY	3400	AGCAAAATCAACCCCACTGAGCTTACCCAGACCGCAGTGCAGCCAGCAGTGTCCAGGTG	3459
Db	3689	AGCAAAATCACTCAACAGAGCCACTTAAGTCACACAGATTAGCGGCGAGTGTCCAGGTG	3748
QY	3460	ATCGGAGTATGCAATTCACACCGCCCAAGACGATGACGAATAGCTTCACAGAAAGGCCAG	3519
Db	3749	ATTGGGATATGACGACTACACCGCCCAAAATGACCATAGCTGGGCTTCAACAAAGGGCCAG	3808
QY	3520	ATCATCAACGTCCTCAACAAGAGAGACCCGGACTGTGTGAAAAAGAGAGTCACTGGGCCAA	3579
Db	3809	ATCATCAACGTCCTCAACAAGAGAGACCCCTGACTGTGTGAAAAGAGAAAGTCAATGAGCA	3868
QY	3580	GTTGGGCTCTTCCCATCCAAATTATGTAAGGTGACCAACAGCATGACCCGACCCAGCA	3639
Db	3869	GTGGGGCTCTTCCCATCCAAATTATGTAAGGTGACCAACAGCATGACCCGACCCAGCA	3928
QY	3640	TG 3641	
Db	3929	TG 3930	

```

RESULT 6
AF064243
LOCUS AF064243 5287 bp mRNA linear PRI 21-NOV-1998
DEFINITION Homo sapiens intersectin short form mRNA, complete cds.
ACCESSION AF064243
VERSION AF064243.1 GI:3859852
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 5287)
Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
Two isoforms of a human intersectin (ITSN) protein are produced by
brain-specific alternative splicing in a stop codon
Genomics 53 (3), 369-376 (1998)
99017974
MEDLINE
9799604
2 (bases 1 to 5287)
Guipponi,M., Scott,H.S., Chen,H., Schebesta,A., Rossier,C. and
Antonarakis,S.E.
Direct Submission
Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland
Location/Qualifiers
1..5287
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107..3769
/codon_start=1
/product="intersectin short form"
/protein_id="AAC78610.1"
/db_xref="GI:3859853"
/translation="MAQFPFPGGSLDIWAIIVERRAKHQQFSLKLPISGFTIGDQA

```


QY	1558	ATTGTGAAATATCCCACTTACAGACAGAGTTGGACGAAATCTCAGCAAAATGCTTGGAGA	1617
Db	1667	ATTGCCGAAATCACCCATCTACAGCAACAATTACAGGAATCTCAGCAAAATCTTGGAGA	1726
QY	1618	CTTATTTCCAGAGAAACAGATATCTCAGTACCCAGTTTAAACAAGTCACAGCAACAGTTTG	1677
Db	1727	CTTATTTCCAGAAAAACAGATATCTCATATACCAATTTTAAACAAGTTCAGCAACAAGTTTG	1786
QY	1678	CATAGAGACTCGCTTCTTAACCTCAAAAAGAGCTTTGAAGCAAAAGAGACTGGCCCGCAG	1737
Db	1787	CACAGAGATTCACCTTGTTACACTTTAAAGAGCTTTAAGAGCAAAAAGACTAGCTGGCAG	1846
QY	1738	CAGCTCCGGGAGCAGCTGACGAGGTGGAGAGAGACCAAGTCAAAGCTCAGAGATTT	1797
Db	1847	CACCTTACAGACCAACTGGAATGAGAAAGAACTAAGATCAAACTACAGAGATTT	1906
QY	1798	GATGTTTCAACAACAGCTGTAAGGAACCTGAGAGAGATATCAATAGCAACAGCAACTCCAG	1857
Db	1907	GATATTTTCAATATACAGCTGAAGGAACCTAAGAGAAATACATATACCAACAACTCCAG	1966
QY	1858	AAGCAGAGTCCCTGGAGGACGCGAGCTGAAGCAGAAAGAGCAGAGAGAGAGAGCCCTG	1917
Db	1967	AAGCAAAAGTCCATGGAAGCTGAAGACTGAACCTGAACAGAAAGAACAGAAAGATCATTA	2026
QY	1918	GAGTTTAGAAGCAAAAGAGAACGCTCAAGACGATTTCAAGAAAGGACCAAGCATGG	1977
Db	2027	GAAATTAGAAAAACAATAAGAAAGAACCCAAAGACAGACTCAGAAAGGGAACAAGCAGTGG	2086
QY	1978	CTGGAGCAATGTCACACAGGAG---GAGAGCCACGCCCCCGGAAACCCACAGAGAGGAC	2034
Db	2087	CTGGAGCAATGTCACACAGAGAGACGAGACTCAGAACCAAGAAACTCAAGAAAGGAA	2146
QY	2035	AGACTGAAGAGGAGAAAGACAGTGTCAAGAAAGAGGCGGGAAGAGAGACCAAGCCGGA	2094
Db	2147	AAACTGAAAAAGGAGAGAGAGTGTCAAAAAGAGATGGCGAGGAAAAAGCAACAGGAA	2206
QY	2095	ATGCAGAACAAGCAGAGTGGCTTTTCAATCCGATCAAGAGCCAGTTAAGTGGCAAC	2154
Db	2207	GCACAAGACAAGCTGGGTGGCTTTTCCATCAACCAAGAACCCAGTAAAGCAGCTGTC	2266
QY	2155	CAGGACCCCTGGTCTTACCAACAGAGAAAGGCCCGCTTACATTTCTGCAACAGAGAGTGA	2214
Db	2267	CAGGACCCCTGGTCTTACCAACAGAAAAAGGTCCACTTACATTTCTGCAACAGAAAAATGA	2326
QY	2215	AAAGTGATATATTTACCGAGCGCTGACCCCTTTGAATCAGAAAGTCAGATAGATCAAC	2274
Db	2327	AAAGTGATATATTACCGGGACTGATACCCCTTTGAATCCAGAAAGCAAGATGAATCACT	2386
QY	2275	ATCCAGCCAGAGATATAGTCAT-----GGTGAAGAAAGCCAGCTGGA	2319
Db	2387	ATCCAGCCAGAGACATAGTCATGGTTAAAGGGGAATGGGTGATGAAGCCAAACTGGA	2446
QY	2320	GAGCCAGAGATGGCTTGGAGAGAGAGCTGAAGGGAGAGAGGATGGTTCCCTGCAAACTAT	2379
Db	2447	GAAACCCGCTGGCTTGGAGAGAAATTTAAAGGAAAGACAGGGTGGTTCCCTGCAAACTAT	2506
QY	2380	GCAGAAAAAGATTCCAGAAAAATGAGGTTCCCACTCCAGCCAAACCAAGTACCGATGTGACA	2439
Db	2507	GCAGAAAAAATCCAGAAAAATAGGTTCCCGCTCCAGTGAAGAACAGTGAATGATTTCAACA	2566
QY	2440	TCGTGCCCCGCCCAAACTGGACTGTGCGTGAAGACCCCTGCTCTTTGCCAGTGAACCTT	2499
Db	2567	TCGTGCCCCGCCCAAACTGGACTTGGGTGAAGACCCCGGCCCTTTGGCAGTAACTCT	2626
QY	2500	TTTGAGCCCTCCAAACACCCCAACAACCTGGGCACTTCAAGTTCAAGTGGCCACGACG	2558
Db	2627	TCAGAGCCCTCCACACACCCCTTAATTAATCTGGGCGCACTTCAGCTCCAGTGGCCCAACG	2688
QY	2560	TCAAACGAGAACCCAGAAACGGAACAATGGGATAAGTGGGCGGCTCAGCTTCTCTGACC	2619
Db	2687	ACGAATGAGAAACCAAGAACGATTAATCGGAGTGAATGGGACGCCACGACCTCTTCAACC	2748
QY	2620	GTACCTTAGTGTGGCCAGTTAAGGAGAGATAGGCTTTTACCCCAAGCACAAGCACTGAC	2679

Db	2747	GTTCAGATGCGGGCCAGTTAAGGCAAGGTCGCCCTTACTCCAGCAGGCCACTGGCC	2806
QY	2680	TCCCTCCCATCTCCCGCTCTGGGCGAGGGTGAAGAGGTGAAAGGGCTTACAAGCCGAAGCC	2739
Db	2807	TCCCTCCCGCTCTCGTGTGCTAGGCGCAAGGGTGAAGAGGTGAGGGGCTTACAAGCTCAAGCC	2866
QY	2740	CTGTATCCCTGGAGAGCCAAAAAGACACACCTTAATTTTAACTAAAGTGCATC	2789
Db	2867	CTATATCTCTGGAGAGCCAAAAAGACACACCTTAATTTTAACTAAAGTGCATC	2926
QY	2800	ACCGTTCGGAACGCAAGACATGTGGTGGTTGGAGAGTTCAAGGTCAAGAGGTGG	2859
Db	2927	ACCGTTCGGAACGCAAGACATGTGGTGGTTGGAGAGTTCAAGGTCAAGAGGTGG	2986
QY	2860	TTCCCAAGTCTTACGTGAACCTATTTCAAGGGCCCGTAAGAAATCACAAGCATCGAT	2919
Db	2987	TTCCCAAGTCTTACGTGAACCTATTTCAAGGGCCCGTAAGAAATCACAAGCATCGAT	3046
QY	2920	ACTGGCCCTACTGAAGCTCTGTGATGTCTAAGAGATGGCTTCCCGCCCGCAAGCCA	2979
Db	3047	TCTGGTCTTACAGAGAGTCTGTGATGTCTAAGAGATGGCTTCCCGCCCGCAAGCCA	3106
QY	2980	GCCATTCGCCGAGAGAGTTATTCGATGATACATACGAGAGTTCTGAGCAAGAGAT	3039
Db	3107	GTCGTTTCGGAGAGAGATTTATGTCATGTACCTTACGAGATTTGAGCAAGAGAT	3166
QY	3040	TTAACCCTTTCAGCAAGGGAGTGTGATTTGGTTTACCAAGAAAGATGTGACTGTGACG	3099
Db	3167	TTAACCCTTTCAGCAAGGGAGTGTGATTTGGTTTACCAAGAAAGATGTGACTGTGACG	3226
QY	3100	GGAACGGTGGCGCAAGTCCGAGTCTTCCCTTCTAATGTGAGCTTTAAAGATTCA	3159
Db	3227	GGAACGGTGGCGCAAGTCCGAGTCTTCCCTTCTAATGTGAGCTTTAAAGATTCA	3286
QY	3160	GAGGGCTCTGGAATGCTGGGGAACAGGGAGTTAGGAAGAAAACTGAAATGGCCAG	3219
Db	3287	GAGGGCTCTGGAATGCTGGGGAACAGGGAGTTAGGAAGAAAACTGAAATGGCCAG	3346
QY	3220	GTTATTTGCTTCTAAGCTGTGCTACTGTGCCGACACTCACTCGTGGTCTCTGGCAGCTG	3279
Db	3347	GTTATTTGCTTCTAAGCTGTGCTACTGTGCCGACACTCACTCGTGGTCTCTGGCAGCTG	3406
QY	3280	ATTCTGATCCGGAAGAAAGAACCCAGGTGATGTGGGAAGGAATCTCAAGCTTCGAGCG	3339
Db	3407	ATTCTGATCCGGAAGAAAGAACCCAGGTGATGTGGGAAGGAATCTCAAGCTTCGAGCG	3466
QY	3340	AAAAAGGCCAGATAGGGTGGTTCCAGGCAATATATGTCAAACTTTAAGCCCGGAACA	3399
Db	3467	AAAAAGGCCAGATAGGGTGGTTCCAGGCAATATATGTCAAACTTTAAGCCCGGAACG	3526
QY	3400	AGCAAAATCAACCCCACTGAGCTACCAAGACCCGACGTGACGACGAGTGTCCAGGTG	3459
Db	3527	AGCAAAATCACTCCAAAGAGCCACTTAATGTCAACACATTAAGGCGCAGTGTCCAGGTG	3586
QY	3460	ATCGGAGTATGATTAACACCGCCGAGACGATGACGAATGAGCTTCCAGAAAGGCGAG	3519
Db	3587	ATCGGAGTATGATTAACACCGCCGAGACGATGAGCTTCCAGAAAGGCGAG	3646
QY	3520	ATCATCAACGCTCCCAAGAGGAGACCCGAGCTGTGGAAAGGAAGTCAAGTGGGCA	3579
Db	3647	ATCATCAACGCTCCCAAGAGGAGACCCGAGCTGTGGAAAGGAAGTCAATGACAA	3706
QY	3580	GTTGGGCTCTTCCATCCATTAATGTAAGCTGACACAGACATGGAACCCGACGACAA	3639
Db	3707	GTTGGGCTCTTCCATCCATTAATGTAAGCTGACACAGACATGGAACCCGACGACAA	3766
QY	3640	TGA 3642	
Db	3767	TGA 3769	

RESULT 7

AF064244
LOCUS AF064244 7247 bp mRNA linear PRI 21-NOV-1998
DEFINITION Homo sapiens intersecin long form mRNA, complete cds.
ACCESSION AF064244
VERSION AF064244.1 GI:3859854
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 7247)
AUTHORS Guipponi, M., Scott, H. S., Chen, H., Schebesta, A., Rosasier, C. and Antonarakis, S. E.
TITLE Two isoforms of a human intersecin (ITSN) protein are produced by brain-specific alternative splicing in a stop codon
JOURNAL Genomics 53 (3), 369-376 (1998)
PUBMED 99017974
FEATURES
PUBMED 9799604
REFERENCE 2 (bases 1 to 7247)
AUTHORS Guipponi, M., Scott, H. S., Chen, H., Schebesta, A., Rosasier, C. and Antonarakis, S. E.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue Michel-Servet, Geneva 4 CH-1211, Switzerland
LOCATION/Qualifiers
1. 7247
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/cisue_type="brain"
/dev_stage="fetus"
107..5272
/codon_start=1
/product="intersecin long form"
/protein_id="AAC78611.1"
/db_xref="GI:3859855"
/translation="MAOPPTREGSLDIAIVTEBAKDDQHSKIPISGITGQA
RNPFRSGIPQVLAQIWMALAMNDGRDQVFSIAMKLIKLOGVLPSPALPVM
KOQPAVISAAPPFGWGLASMPPLTAVAPMGSIIPVMSPLTSSVTPAVPLAN
GAPVLOPLPAPFAPPAATLPSKSSFSRSGSOLNTLKAQSFDAVAPVEMAVP
OSRLKRYOLFNSHDMGSLTGPQARTLMOSLIPOLASINLSIDIDQKULTA
EEFLIAMHLIDVAMSGOPLPVLPPEVTPSPRRVSGSGISVISTSYDOLPEPV
LEBQOOLEKLPVTFEDKRENFEGNILEKROALLBOQKEBERLAQUBRAOE
KREBERQBERKOLEKOLEKOREBERKEREKEREKEREKEREKEREKEREKERE
RNRQELNQRNKEODIYVLKAKKTLFEELANDKHOLEGKLODJRCRLTTORO
EIESTKSELELAETLHLOQLOESQOMGLIPEKQILNDLOKVOONSLSHRLSV
TLKRALEKELAROHRLDOLDEVERETRSLOEIDIENOLKELEIHHKOOLOKKS
MEARERKOEKOEKIELEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEK
KAEESYKODGEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEK
KVAYTALYFESRSHDEITIQGDIIVMKGWVDSQGEPMGLKGLKGTGEPFA
NNAEKIPENEVAPVAVPVDTSAPAKLARETAPLAVTSSEPTNNMNADEFT
WPTSTKEPETNDMAAOPSLTVPASQLOKRSAPFATATVSEPTNNMNADEFT
GQAQALYEMRAKDNHNLNNKNDVITVLEODMWEVGOGKGPSPYLSGPEVE
IKRSTMDSGSSSPASLKRVAAPKPVASGEPIAMTVTSSBOGDI.TPOGQDYL
VTKKDDMMTGYTGDKAGVPSPSYVALKSSBSGTGKTSGLKKEPIAOVASTAT
GPEQLTLAEGQLLIRKNKPGWEGELDARGGKROIGMPAYVALNLNPKSTPT
EPKSTALAAVOCVIMDYTAQNDDELAFNKQIINVANKEDPMWKEBVGVOGLF
PSNYVLTITDMDPSQOMCSDLHLDMITPERKQGYIHLLIVTEBNYVNDLOVTEI
FOKPLMESELDEKEVAMI.FVMKELIMCNILKALRVKMKSGKMKVMKMGIDLS
AOLPHMOPYIRFCSROLNGAALIOCKTDAPEKFEVRLKLEMDPRCKMPLSSTLKP
MORVTRPLIKIKIILENTBNHDSHLKALEKAEELCSQVNEGVREKENSRLLEMT
QAVOCEGLSEQLVENSVTNCJGPRKFLHSGKLYKAKNKELYGFENFLITLOITK
PLGSGTDVFPSPKSNLYKMYKTPILFLEBVVLKPLDPSGDEPIFIHISIVRYLR
AESINERTAMVOKIKAASELVLETEKKEKAVLVSQATGIGRLVAVVSDIEKLP
CSHSGSNPYCEVTGMSOCHITKTIOTDLPKNMNSCOPIRLBOEVLCITYFEEDQ
ESPDDFLGRTETIRVADIKKQGSKGPVTKCLLHREVPTEIYVRDLQLEPDR"
167..406
/note="encodes EH domain"
767..1936
/note="encodes EH domain"
2324..2524
misc_feature
misc_feature
misc_feature

/note="encodes SH3 domain"
2843..3019
/note="encodes SH3 domain"
3110..3286
/note="encodes SH3 domain"
3326..3520
/note="encodes SH3 domain"
3569..3748
/note="encodes SH3 domain"
3836..4390
/note="encodes SH3 domain"
4649..4819
/note="encodes SH3 domain"
4895..5143
/note="encodes SH3 domain"
2156 a 1652 c 1747 g 1692 t
BASE COUNT
ORIGIN
Query Match 77.6%; Score 2826.8; DB 9; Length 7247;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 3164; Conservative 0; Mismatches 477; Indels 21; Gaps 3;
QY 1 ATGGCTAGTTCCACACCTTCGCTGAGCTGATGTCGGCCATACTGAG 60
DB 107 ATGGCTAGTTCCACACCTTCGCTGAGCTGATGTCGGCCATACTGAG 166
QY 61 GAAAGGCCAAGCATGACGACGATTCCTTACCTGAGCCGATGAGGATTTACT 120
DB 167 GAAAGGCCAAGCATGATGACGATTCCTTACCTGAGCCGATGAGGATTTACT 226
QY 121 GGTATCAAGCAGGAACTTTTTCATCTGGGTTACTCAGGCTGCTTAGACAA 180
DB 227 GGTATCAAGCAGGAACTTTTTCATCTGGGTTACTCAGGCTGCTTAGACAA 286
QY 181 ATATGGCGCTAGCGGACATGATTAAGATGAGATGATCAAGTATTTCCATA 240
DB 287 ATATGGCGCTAGCGGACATGATTAAGATGAGATGATCAAGTATTTCCATA 346
QY 241 GCATGAGACTTATCAACTGAAGTACAGGATATCAGCTCCCTCCACATTCCTCC 300
DB 347 GCATGAGACTTATCAACTGAAGTACAGGATATCAGCTCCCTCCACATTCCTCC 406
QY 301 GTCATGAACACGACACAGTGGCTATTTCCATGACGACGACGATTTGATAGAGGATT 360
DB 407 GTCATGAACACGACACAGTGGCTATTTCCATGACGACGACGATTTGATAGAGGATT 466
QY 361 GCTAGATCCACACCTCAGCTGTGCTCCTGTCAGATGAGGCTCATTCAGTTGTT 420
DB 467 GCTAGATCCACACCTCAGCTGTGCTCCTGTCAGATGAGGCTCATTCAGTTGTT 526
QY 421 GGAATGCTCCACCTTATGATCTTGTCTCCTCAGACGAGTGTCTCCCTGCTAAC 480
DB 527 GGAATGCTCCACCTTATGATCTTGTCTCCTCAGACGAGTGTCTCCCTGCTAAC 586
QY 481 GGGGCTCCCTCGGTATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 587 GGGGCTCCCTCGGTATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 646
QY 541 AAGAGTCTCTCTGAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 647 AAGAGTCTCTCTGAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706
QY 601 GCAATATCTGATGTCGCGACGCGCTCCAGACGACGAAATGGGCTGCTGCTGCTGCT 660
DB 707 GCAATATCTGATGTCGCGACGCGCTCCAGACGACGAAATGGGCTGCTGCTGCTGCT 766
QY 661 TCAAGCTGAATTAAGAGAGTATTAACACGACGACGAAATGAATGAGTGAACACTTA 720
DB 767 TCAAGCTGAATTAAGAGAGTATTAACACGACGACGAAATGAATGAGTGAACACTTA 826
QY 721 ACAGGTCCTCCAGCAGCAAGTATTCATGATGATGATGATGATGATGATGATGATGAT 780
DB 827 ACAGGTCCTCCAGCAGCAAGTATTCATGATGATGATGATGATGATGATGATGATGAT 886

QY	781	CCAAATATGGAATCTTTTCGACATGATCAAGATGGAAAACTCACATGCAAGAAATTTATC	840
Db	887	TCAAATATGGAATCTTTTCGACATGATCAAGATGGAAAACTTACACGAGGAATTTATC	946
QY	841	CTAGCTATGACACTTAATGATGTTGTCAGTGTGATCGACCACTGCGCGCTCTGCGCT	900
Db	947	CTGGCAATGACCTCATGATGATGATGCTGAGCCAACTGCGCACTGTCTCTGCGCT	1006
QY	901	CCAGAAATACATCCCTCTCTCTTCAGAAAGATTCGCTCCGGAGTGGGATGTCCGTATA	960
Db	1007	CCAGAAATACATCCCTCTCTCTTCAGAAAGATTCGCTCCGGAGTGGGATGTCTGTATATA	1066
QY	961	AGCTTTTCTTCTGTGATCAGAGGTGCTGAGAGCCGTCGTCAAGAGATGACAGAG	1020
Db	1067	AGCTCAACATCTGTGATCAGAGGTGCTCAGAGAGAACAGTTTGAAGATGACACATA	1120
QY	1021	C--CAGAGAGAACTGCTGTGACATTTGAAGTAAAGCCGGAAGACTTCAGACGA	1077
Db	1127	CAATTAGAAAAAGAAATTAACCTGTACGTTTGAAGTAAAGCCGGAAGACTTCGAACGT	1188
QY	1078	GGCACATGAGAGCTGGAAGAGCGGCCAAGGCTCTTGAGACACAGCGCAAGACAG	1133
Db	1187	GGCACATCTGGAATCTGAGAAAGCAAGCTCTCTGGAACACACAGCGCAAGACAG	1246
QY	1138	GAGCGGTGGCTCACCTGAGCGCCCGACGACGAGAGAGAAAGACGGGAGCGCCAGAG	1197
Db	1247	GAGCGCTGGCCCACTGAGCGCGCGGAGCAAGAGAGAGACGTGAGCGCCAGAG	1306
QY	1138	CAGAGAGCCAAAGCCGAGCTGGAAGCTGGAAGACGCTGGAAGACAGCGGAGCTGGAG	125
Db	1307	CAGAGAGCCCAAGCAACTGGAACCTGGAAGACACTGGAAGAACAGCGGAGCTTAA	1366
QY	1258	CGGACGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAACGGGAA	1317
Db	1367	CGGACGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAACGGGAA	1422
QY	1318	CTGGAAGGACGCGACAACTTGATGGAGCGGAACCGGAGACAGAACTCTGTATCAG	1377
Db	1437	CTTGAAGGACGCGACAACTTGATGGAGCGGAATCGGAAGCAAACTATAATCAA	1486
QY	1378	AGGAACAAGGAGACAGAGGCGACGCTGCTGGAAGGCAAGGAGGAAGACTCTGAGTTT	1433
Db	1487	AGGAACAAGGAGACAGAGGCGACTGTTGATCTGAAGCAAGAAAGACTTTGGAATTT	1546
QY	1438	GAGTTAGAACTCTGATATGACAAAAAGCATCAGCTTAGAAGGAAAACTTCAGATATCAG	1497
Db	1547	GAAATTAGAAAGCTCTAATATGATATAAAGCATCAACTAGAAGGAACTTCAGATATCAG	1606
QY	1438	TGTCGACTGGCAACCCAGAGGCAAGAAATTGAGAGACGAACAACTTAGAGAGCTAAGA	1557
Db	1607	TGTCGACTTGGCAACCCAAAGGCAAGAAATTGAGAGACGAACAACTTAGAGAGTTGAGA	1666
QY	1558	ATTGCTGAAATCACCCCACTTACAGGACGTTGCAAGAACTTCACAAATGCTTGAAGA	1617
Db	1667	ATTGCTGAAATCACCCCATTTACAGCACAATTCACAGAACTTCACAAATGCTTGAAGA	1722
QY	1618	CTTATTCAGAGAAACAGATACTCAGTACGACGTTTAAACAAAGTCACAGAACAGTTTG	1677
Db	1727	CTTATTCAGAGAAACAGATACTCAGTACGACCAATTAAACAAAGTTTCAGAGAACAGTTTG	1786
QY	1678	CATTAAGACTGCTCTTACCTCAAAAGAGCTTGGAAGCAAGAGCTGCGCGGAG	1733
Db	1787	CACAGAGATTCACCTGTTCACCTTAAAGAGCTTGAAGCAAGAACTTACCTGGCAG	1846
QY	1738	CAGCTCCGGAGACAGCTGAGCCAGGTGAGAGAGAGACCAGGTCAAAGCTGACAGAGTT	1797
Db	1847	CACCTACAGAGCAACCTGAGTGAAGTGAAGAAACCTAAGATCAAACTTACAGAGATTT	1906
QY	1738	GATGTTTTCACACACAGCTGAAGGATCTGAGAGATTCATACGAACACGACTCCAG	1857
Db	1907	GATATTTTCAATTAATCAGCTGAAGGATCTTAAGAGAAATTCACAAATTAAGCAACCTCCAG	1966

OY	1858	AAGCAGAGGTCCTCGAGAGCGAGCCGCGATGAAAGCAGAAAGACGAGAGAGGAAAGAGCTG	1917
Db	1867	AAGCAAAAGTCCATGAGAGGCTGAACGACTGAACAGAAAGAACAGAAACGAAATCATTA	2026
OY	1918	GAGTTAGAGAAACAAAGAAAGCGCTCAGAGACGAGATTCCAGAAAGGACCAAGCAATGG	1977
Db	2027	GAATTAGAAAAACAAAAGAAAGAACCCAAAGACGACTCAGAAAGGAGACAGCACTGG	2086
OY	1978	CTGGAGCATGTGCAGCAGAG--GAGCAGCCAGCCCCCGGAAACCCACGAGAGAGAC	2034
Db	2087	CTGGAGCATGTGCAGCAGAGAGCAGACATCAGAGACCAAGAAATCTCACAGAGAGAA	2146
OY	2035	AGACTGAAGAGGAGACAGCTGTGAGGAAGAAAGGAGCGGAAAGAGAGAGCCAGCCGAA	2094
Db	2147	AAACTGAAAAGGAGAGAGTGTCAAAAAGAGATGCGAGGAAAAAGGCAACGAGAA	2206
OY	2095	ATGCAGACAAAGCAGAGTCGGCTTTTCATCCGATCAGAGAGCAGCTAAGCTGGCAC	2154
Db	2207	GCACAGACAGCTGGGTCGGCTTTTCATCAACACAAAGAACAGCTAAGCCAGCTGT	2266
OY	2155	CAGGACCCCTGGTCTACACAGAGAAAGGCCCGCTTACATTTCTGACAGAGAGTGA	2214
Db	2267	CAGGACCCCTGGTCCATCGAGAAAAGGTCACTTACATTTCTGACAGAGAAATGTA	2326
OY	2215	AAAGTGTATTTACCGAGCGCTTACCCCTTTGAATCCAGAACTCAGATGAGATCAC	2274
Db	2327	AAAGTGTGTATTACCGGAGCTGTATCCCTTTGAATCCAGAAACCATGAGAAATCACT	2386
OY	2275	ATCCAGCCAGAGATATATGTCAT-----GGTGAATGAAGCCAGACTGGA	2319
Db	2387	ATCCAGCCAGAGACATATGTCATGTTAAAGGGAATGGGTGATGAAGCCAACTGGA	2446
OY	2320	GAGCCAGAGATGGCTTGGAGAGAGACTGAAGGGAAGACGGGATGGTTCCCTGCAACTAT	2379
Db	2447	GAACCCGCGTGGCTTGGAGAGATTAAGGAAGACAGGGTGGTTCCCTGCAACTAT	2506
OY	2380	GCAGAAAAAGATCCAGAAAATGAGTTCCACTCCAGCCAAACAGTGAACCGATCTGCA	2439
Db	2507	GCAGAGAAAATCCAGAAAATGAGTTCCCGCTCAGTGAACCAAGTACATTCATCA	2566
OY	2440	TCTGCCCTCGCCCCCAACTGGGCTCTCGGTGAGACCCCTGCTCTTTGCCAGTGAACCT	2499
Db	2567	TCTGCCCTCGCCCCCAACTGGGCTCTGGGTGAGACCCCGCTCTTTGGCAGTAACTCT	2626
OY	2500	TCTGAGCCCTCCAAACCCCAACCACTGGGCACTTCAGTTCCAGTGGCCGACAGAC	2559
Db	2627	TCAAGGCCCTCCAGACCCCTTAATACGTGGCCGACTTCAGCTCAGTGGCCACAGAC	2686
OY	2560	TCAACAGAAAGCCAGAAACGAGCAACTGGGATACGTGGCGGCTCAGCTTCTCTGACC	2619
Db	2687	ACGAAATGAAGAACAGAAACGAGATACGTGGATCAGTGGCAGCCAGCCCTCTCTCAC	2746
OY	2620	GTACTTATGCTGGCCAGTTACCGCAGAGATCAGCCTTTAACCCAGCCACAGCCACTGGC	2679
Db	2747	GTTCCAAGTGCAGGACAGTTAAGGCAAGGTCCGCTTTATCTCAGCCACGGCCTAGGC	2806
OY	2680	TCTTCCCATCTCCCGTCTGGGCGCAGGAGTGAAGAGGAGGCTTAAAGCCCAAGCC	2739
Db	2807	TCTTCCCGCTTCTCTGTGCTAGGCGAGGATGAAGAGTGAAGGAGGCTTAAAGCTCAAGCC	2866
OY	2740	CTGTATCCCTGAGAGCCAAAAAGAACCACTTAATTTTAACAAAAGTGAAGTCAATC	2799
Db	2867	CTAATATCCTTGGAGAGCCAAAAAGAACCACTTAATTTTAACAAAATGATGTATC	2926
OY	2800	ACCGTTCTGGAACGCAAGACATGTGGTTTGGAGAGTTCAAGTCAAGAGGTTGG	2859
Db	2927	ACCGTCTCGGAACGCAAGACATGTGGTTTGGAGAGTTTCAAGTCAAGAGGTTGG	2986
OY	2860	TTTCCCAAGTCTTACGGAACCTAATTTACGGGCGCGTAAGGAATCAACAAGCATGAT	2919
Db	2987	TTTCCCAAGTCTTACGGAACCTAATTTACGGGCGCGTAAGGAAGTCTACACAAGCATGAT	3046
OY	2920	ACTGGCCCTACTGAAGTCTCTGTAGTCTAAGAGAGTGGCTTCCCGGCGCCAGGCA	2979

QY	421	GGAAATGTCTCCACCCCTTAGATATCTTCTGTCCCTCCAGACAGAGTGCCTCCCTGGCTAAC	480
Db	461	GGAAATGTCTCCGCCCTTAGATATCTTCTGTCCCTCCAGACAGAGTGCCTCCCTGGCTAAC	520
QY	481	GGGGTCTCCCGGTCAATACAGGCTCTGCCTGGCTTGGCAATCCTGCAGGCCATAGGSCA	540
Db	521	GGGGCCCTCCGTATATACAGCTCTTGCTGATTTGCTCATCTCTGCAGCCCAATTGGCA	580
QY	541	AAGATTCTTCCCTCAGACAGATCTGGTCCAGGGTCACAATTAAACATTAAGTTACAGAA	600
Db	581	AAGATTCTTCCCTCAGACAGATCTGGTCCAGGGTCACAATTAAACATTAAGTTACAGAA	640
QY	601	GCACATCATTTGATGTGTGGCCGCCCTCCAGACAGAAATGGGCTGTGCTCAATGCA	660
Db	641	GCACATCATTTGATGTGTGGCCGCCCTCCAGACAGAAATGGGCTGTGCTCAATGCA	700
QY	661	TCAAGGCTGAAATATACAGGAGTTATTCAACAGCCACGACAAATCATAGTGTGACACTTA	720
Db	701	TCAAGACTGAAATATACAGGAGTTATTCAACAGCTACGACAAACATATAGTGTGACACTTA	760
QY	721	ACAGGTCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGCT	780
Db	761	ACAGGTCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGCT	820
QY	781	TCGAATATGGAATCTTTCTGCATCTTGATCAAGATGGAATACTACTGCAAGAAATTTATC	840
Db	821	TCGAATCTGGAATCTTTCCGACATTTATCAAGATGGAATACTACCGCAAGAAATTTCAATC	880
QY	841	CTAGCTATGACCTTAATTGATTTGGTGCATGTGGTCAAGCCATGCGCGGCTCCTGCT	900
Db	881	CTAGCGATGACCTTAATTGATTTGGTGCATGTGGTCAAGCCATGCGCGGCTCCTGCT	940
QY	901	CCAGAATACATCCCTCTCTCTTCCAGAAAGATTGCTCCGCGAGTGGATGTCCGTATTA	960
Db	941	CCAGAATACATCCCTCTCTCTTCCAGAAAGATTGCTCCGCGAGTGGATGTCCGTATTA	1000
QY	961	AGCTCTTCTTCTGTGATCAGAGGCTGCTTGAGAGCCGTGTCAGAGATGACAGCAG	1020
Db	1001	AGCTCTTCTTCTGTGATCAGAGGCTGCTTGAGAGCCGTGTCAGAGATGACAGCAG	1060
QY	1021	CCAGAGAAAGAACTGCTGTGATCATTTGAAAGATTAAGACCGGGAACCTTCGAGGAGGC	1080
Db	1061	GTGGAAGAAAGAACTGCTGTGATCATTTGAAAGATTAAGACCGGGAACCTTCGAGGAGGC	1120
QY	1081	AGTGTGAGCTTGGAGAAAGCGCCGCAAGCGCTCTTGAGACAGCAGCAAGAGCAGAG	1140
Db	1121	AACGTGAGCTTGGAGAAAGCGCGCCCTCTGAGACAGCAGCAAGAGCAGCAGAG	1180
QY	1141	CGGTTGGCTCAGCTGAGCGCGCCGACAGAGAGAGAAAGCGGAGCGCCAGAGCAG	1200
Db	1181	CGGCTGGCTCAGCTGAGCGCGCCGACAGAGAGAGAAAGCGGAGCGCCAGAGCAG	1240
QY	1201	GAGGCGAAGCGGCTGAGAGCTGAGAAAGAGCTGAGAAACAGCGGGAGCTGAGAGGG	1260
Db	1241	GAGGCGAAGAGCAAGCTTGAAGCTTGAAGAGCAGCTTGAAGAAACAGCGGGAGCTGAGAGGG	1300
QY	1261	CAGCGAGAGAGAGAGAGAGAGATCGAGGCGCGAAGCGCAAAACGGGAACTG	1320
Db	1301	CAGAGAGAGAGAGAGAGAGAGAGATCGAGGCGCGAAGCGCAAAACGGGAACTG	1360
QY	1321	GAAAGCAGCGCAACTTGAATGGGAACGGAACCGGAGCAGGAATCTCTGAATCAGAG	1380
Db	1361	GAAACCGCAGCGCAACTTGAATGGGAACGGAACCGGAGCAGGAATCTCTGAATCAGAGG	1420
QY	1381	AACAGAGAGAGAGAGAGCGCTGTGCTCTGAAGAGCAAGAGGAACATCTTGAAGTTTGA	1440
Db	1421	AACAGAGAGAGAGAGAGCGCATCTGTGTTTGAAGAGCAAGAGGAACATCTTGAAGTTTGA	1480
QY	1441	TTAGAGAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGAAACTTCAGGATTCAGGTGT	1500
Db	1481	TTAGAGAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGAAACTTCAGGATTCAGGTGT	1540
QY	1501	CGACTGGCAACCCAGAGCCAGAAATGAGACAGCAAGCAAGTCTTAGAGAGCTTAAGATT	1560

Db	1541	CGACTGGCAACCCAGAGGCAAGAAATTGAAGCAAAACAAAGTCTTAAGAGCTGAGAAATT	1600
Qy	1561	GCTGAAATTCACCCACTTACAGCAGCAGAGTTCAGAGAACTCAGCAAAATGCTTGGAAAGCTT	1620
Db	1601	GGCGAATTCACCCACTTACAGCAACAACTTCGAGGAATCTCAGCAGAAATGCTTGGAAAGCTT	1660
Qy	1621	ATTCGAGAAACAGATATCTCAGTGCACCAAGTTAAACAACTCCAGCGCAACAGTTTGCT	1680
Db	1661	ATTCGAGAAACAGATATCTCAGTGCACCAAGTTAAACAACTCCAGCGCAACAGTTTGCT	1720
Qy	1681	AGAGACTCGCTCTTACCCCTCAAAAGAGCCTTGGAAACAAAGAGAGCTGGCCCGGACAG	1740
Db	1721	AGAGATTCGCTCTTACCCCTCAAAAGAGCCTTGGAAACAAAGAGAGCTGGCCCGGACAG	1780
Qy	1741	CTCCGGAGCAGCTGAGCGAGGTGGAGAGAGACCAAGTCAAGCTTCAGAGAGATTGAT	1800
Db	1781	CTTGAGAGGAGCGTGGACGAGGTGGAGAAAGAGACCAAGTCAAGCTTCAGAGAGATTGAT	1840
Qy	1801	GTTTTCAACAAACAGCTGAAGAGAACTGAGAGAGATCATATGCAAAACGCACTTCAGAG	1860
Db	1841	GTTTTCAACAAACAGCTGAAGAGAACTGAGAGAGATCATAGTAAACAGCAGCTTCAAG	1900
Qy	1861	CAGAGGTCCCTGGAGGAGCGCGCATGAGCGAGAAACAGCAGGAGAGAAAGCCTGGAG	1920
Db	1901	CAGAGGTCCATCGAGGCGGAGGCTAAGCAGAAAGCAGGAGAGAGAAAGCCTGGAG	1960
Qy	1921	TTAGAGAAAGCAAAAGAAAGCGCTCAGAGACGAGTTTCAGAAAGGGAACAAGCAATGCTG	1980
Db	1961	TTGGAGAAAGCAAAAGAAAGAGTCAAGAGACGATTCAGAAAGGGAACAAGCAATGCGAG	2020
Qy	1981	GAGCATGTGCAGCAGAGAGAGCAGCCACGCCCTCCGAAACCCCAAGAGAGAGACAGACTG	2040
Db	2021	GAGCATGTCCAGCAGAGAGAGCAGCAGCGCTTCGAGAAACCCCAAGAGAGAGACAAACTG	2080
Qy	2041	AAGAGGGAAGCAATGTCAGGAAABAAGAGCGGGAAGAAGAGAGCCGAGAAATGCCAA	2100
Db	2081	AAAAGGGAAGCAATGTCAGGAAABAAGAGCGGGAAGAAGAGAGCCGAGAAATGCCAA	2140
Qy	2101	GACAGAGAGAGTCGGCTTTTTCATCCGCATTCAGAGCAGAGCTTAAGTGGCCACCCAGCA	2160
Db	2141	GACAGAGAGAGTCGGCTTTTTCATCCATCACAATCAGAGCAGAGTAAAGCCGG---CCGAGCA	2197
Qy	2161	CCCTGTCTACACAGAGAAAGGCCCGCTTACATTTCTGCAAGAGAGTGTAAAGTG	2220
Db	2198	CCCTGTGCCACAGACAGAGAAAGTCCGCTTACATCTCTGCAAGAGAGTGCMAAAGTG	2257
Qy	2221	GTAATATACAGAGGCGCTGTACCCCTTTGAATCCAGAAAGTCAAGATGAGATCACATCCAG	2280
Db	2258	GTAATATACAGAGGCGCTGTACCCCTTTGAAGTCCAGAAAGTCAAGAGATCACATCCAG	2317
Qy	2281	CCAGAGATATATAGTCAT-----GGTGGATGAAAGCCAGACTGAGAGACCA	2325
Db	2318	CCAGAGATATATAGTCATGTTTAAAGGGGAATAGGTGGATGAAAGCCAGACCGAGAGACCA	2377
Qy	2336	GGATGCTTGGAGAGAGCTGAAAGGGAAGCGGAGTGTTCCTGCAAACTATATGCAAA	2385
Db	2378	GGATGCTTGGAGAGAGACCGAAGGGGAAGACAGATGTTCCTGCAAACTATATGCAAG	2437
Qy	2438	AAAGATTCAGAAATGAGATTCACATCCACATCCAGCCAAACCAAGGACCAATGTACATGTGCC	2497
Db	2446	CCTGCCCCCAAACTGGCTCTGCGTGAAGCCCTGCTCCTTTGGCAGAGTACCTTCTTGAG	2505
Qy	2498	CTTGCCCCCAAACTGGCTCTGCGTGAAGCCCTGCTCCTTTGGCAGATGACCTTCTTGAG	2557
Db	2506	CCCTTCACAAACCCCAACAACTGGGAGACTTCAGTTCCAGTGGCCACAGAGCTCAAC	2565
Qy	2558	CCCTTCACAAACCCCAACAACTGGGAGACTTCAGTTCCAGTGGCCACAGAGCACAAT	2617
Db	2566	GAGAAAGCAAAACGAGACACTGGGATACGTGGGCGGCTCAGGCTTCTGTACCGTACT	2625

Db 2618 GAGAAACGAAACGAGCACTGGGACAGCTGGCGCTCAGCTTCTCTGACTGTACCC 2677
QY AGTGTGGCCAGTAAAGGAGAGATCAGCTTATCCCGAGCCAGCAGCACTGGCTCTCC 2685
Db 2678 AGTGTGGCCAGTAAAGGAGAGATCAGCTTATCCCGAGCCAGCAGCACTGGCTCTCC 2737
QY 2686 CCATCTCCCGTCTGGGCGCAGGCTGAAAGGTGAAAGGCTCAAGCGCAAGCCCTGTAT 2745
Db 2738 CTTTCTCCCGTCTGGGCGCAGGCTGAAAGGTGAAAGGCTCAAGCAAGCCCTGTAT 2797
QY 2746 CCTTGAAGAGCCAAAGAAAGCAACCTTAATTTTAACAAAGTACGCTCATCCGTT 2805
Db 2798 CCTTGAAGAGCCAAAGAAAGCAACCTTAATTTTAACAAAGTATGTCATCCGTT 2857
QY 2806 CTGGAACAGCAAGACATGTGTGTGTTTGAAGAGTTCAAGGTGAAAGGTTGGTCC 2865
Db 2858 CTGGAACAGCAAGACATGTGTGTGTTTGAAGAGTTCAAGGTGAAAGGTTGGTCC 2917
QY 2866 AAGTCTTACGTGAACTCAATTTTCAGGGCCGTAAGAAATCCACAAGCATCATCTGAC 2925
Db 2918 AAGTCTTACGTGAACTCAATTTTCAGGGCCGTAAGAAATCCACAAGCATCATCTGAC 2977
QY 2926 CTTACTGAAAGTCTCTGTAGTCTTAAAGAGAGTGGCTTCCCGGCGCCAAAGCCGCTAT 2985
Db 2978 CTTACTGAAAGTCTCTGTAGTCTTAAAGAGAGTGGCTTCCCGGCGCCAAAGCCGCTAT 3037
QY 2986 CCCGGAAGAGATTATTGTCATGTACATACGAGAGTTCTGAGCAAGAGATTAAACC 3045
Db 3038 CTTGGAAG-- 3046
QY 3046 TTTTCAGCAAGGGATGTGATTGTGTTACCAAGAAAGATGTGATGTGACGGAAG 3105
Db 3047 ----- 3046
QY 3106 GTGGCGCAAGTCCGAGTCTTCCCTTCTAATGTGAGGCTTAAAGATTCAAGAGGC 3165
Db 3047 ----- 3046
QY 3166 TCTGGAAGTCTGGGAAACAGGAGTTTAGAAGAAACCTGAATTGCGCAAGTTATT 3225
Db 3047 ----- 3064
QY 3226 GCTTCTACGCTGCTACTGTGTCGGAACAACCTACCCCTGGCTCTGGGAGCTGATTTCTG 3285
Db 3065 GCTTCTACGCTGCTACTGTGTCGGAACAACCTACCCCTGGCTCTGGGAGCTGATTTCTG 3124
QY 3286 ATTCGGAAGAAAGCCAGGTGATGTGGAAGAGAACTGCAAGCTCGAGGAAAG 3345
Db 3125 ATTCGGAAGAAAGCCAGGTGATGTGGAAGAGAACTGCAAGCTCGAGGAAAG 3184
QY 3346 CCGCAGATAGGCTGTTTCCAGCAATTTATGTCAAACTTCTAAGCCCGGAAACAGCAAA 3405
Db 3185 CCGCAGATAGGCTGTTTCCAGCAATTTATGTCAAACTTCTAAGCCCGGAAACAGCAAA 3244
QY 3406 ATCAACCCCAATGAGCTACCCAAAGACCGAGTGCAGCTGATGATCGGAG 3465
Db 3245 ATCAACCCCAATGAGCTACCCAAAGACCGAGTGCAGCTGATGATCGGAG 3304
QY 3466 ATGTACGATTAACCGCGCAAGAACGATGACGAACTAGCCTTCAGCAAGAGCCAGATCATC 3525
Db 3305 ATGTACGATTAACCGCGCAAGAACGATGACGAACTAGCCTTCAGCAAGAGCCAGATCATC 3364
QY 3526 AAGCTCTCAACAAGAGAGACCCGAGCTGTGTGAAGAGAGAGTCACTGTGGCAAGTTGGG 3585
Db 3365 AAGCTCTCAACAAGAGAGACCCGAGCTGTGTGAAGAGAGAGTCACTGTGGCAAGTTGGG 3424
QY 3586 CTTTCTCCCAATCAATTATGTAAAGCTGACCAAGACATGAGACCCGAGCCAGCAATGA 3642
Db 3425 CTTTCTCCCAATCAATTATGTAAAGCTGACCAAGACATGAGACCCGAGCCAGCAATGA 3481

RESULT 9
AF032118

LOCUS AF032118 4103 bp mRNA linear VRT 13-NOV-1998
DEFINITION Xenopus laevis intersecin mRNA, complete cds.
ACCESSION AF032118
VERSION 1 GI:2642624
KEYWORDS
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 4103)
Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P. S., Castagnoli, L., Cesareni, G. and Kay, B. K.
Intersecin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains
J. Biol. Chem. 273 (47), 31401-31407 (1998)
MEDLINE 99030416
JOURNAL 9813051
PUBMED 2 (bases 1 to 4103)
Hoffman, N.G., Hardison, N.L., Yamabhai, M. and Kay, B. K.
Direct Submision
Submitted (30-Oct-1997) Pharmacology, University of Wisconsin, 1300 University Ave, Madison, WI 53706-1532, USA
LOCATION/Qualifiers
1. 4103
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/cell_type="oocyte"
193. 4005
/function="involved in endocytosis"
/note="EH domain and SH3 domain containing protein; similar to intersecin binding proteins Ibp1, encoded by Genbank Accession Number AF057285, and Ibp2, encoded by Genbank Accession Number AF057286, and mouse Rab/Rip, encoded by Genbank Accession Number AF057287"
/codon_start=1
/product="intersecin"
/protein_id="AAC73068.1"
/db_xref="GI:2642625"
/translation="MAQGTPEGNDIMDIWATVERAKHDQFHGKLPAGYITGDOA RNFPLQSLPQVLAQIWLADNMNDGMDQLEFSLAMKILIKLQGYLPSPISNM LKQVAMPAAVAAGMSGIVGIPPLAAVAPVPMSPVSGSPVSPVPPVPLSN GAPAVIQHPRAHSAATLPKSSRBSVAAGSINIKLKQAGSFVDPAFVYEMAVPS SRLRYRQFNSQDKTMSGNLTGPARTILMOSLIPQSLATIMNSIDIDQKLTAS EFLIAMLDVAMSGQPLPPLIPPEYIPSPFRVSGSLSTMSVSDQRLPEEER BEPQNAADKLLPVTFEDKKRBNFERGNLELEKROALIEQOREORLAQLEBAQERK EBERODERKROQDLEKOLEKORLEBERREKRELEBAKELEROLEMERN RROELNORNRQEDIVLKKAKKTLFELALNKKIQLEKLDIRCLRTTORHEI ESTNSRELRALIEITHLQOQLESQILGKMPKQSLIDQKQVQNSLHSDSLTL KSAITKEIGRQQLDQDEVEKETRAKQLEIDVNNQLEKLELNNKQFQKQDFE TEKIQKELERKTSLELDLKEEDKRMLEQDLMDRVKQEBERYKFDQEEKREES YQKCEVKEKPEIEQENKPFQPPPEPGIGQIPMMNTEKALITINQGVKVVYAL YPFDARSHDEITIEPGDILIMVDESQTEPGMLGELKGTGMFPANVAREMESPEPS TTKPAETTKPTVAVAPSPAFAFTNTSTNSNMWADSSSTPNTNMDKVSNDMDT WAGPSLTVPSAGCRQGSATPAPVATVSSSPVUGQGEKVEGLQALYPRAKDN HLNENKNDVITVLEQDMMWGEVQGGKMFPSYVKLISGLRSTSIDSTSSSPA SLKRVSSPAFKPALIQELEYISMTYTESNEQDLTQOQDLIVIKKDGMMWTGYEER TGVPSPNVYRPMQDSAGSGKTSGLKGPKEIAQVYIASAALAPQLTLAPQLILIR KNPQGMWEGELQARKKRQIGMFPANVYKLLSGPTNKSTPEPRKPSLPTQCVIS WYDYIAONDELAFSGKQVINYANKEDDMWGEINLGEFSPSYVVLITLMDSSQO FRUGVPRAGTATGDRPILFPFDGSLPLNAPQAPLPSVMTKFCFTAPRCPD MNVKTINI"
253. 498
/note="encodes Eha domain"
850. 1101
/note="encodes EHB domain"
2386. 2571
/note="encodes SH3A domain"
2881. 3057
/note="encodes SH3B domain"
3148. 3324
/note="encodes SH3C domain"

misc_feature 3364..3558 /note="encodes SH3D domain"
 misc_feature 3607..3786 /note="encodes SH3E domain"
 BASE COUNT 1230 a 920 c 1029 g 856 t 8 others
 ORIGIN

Query Match 55.7%; Score 2027.6; DB 5; Length 4103;
 Best Local Similarity 73.6%; Pred. No. 0; Mismatches 918; Indels 39; Gaps 7;
 Matches 2685; Conservative 4;

1 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCCTGTAGTGTCTGGGCCATTACTGTGAG 60
 Db ATGGCTCAGTTTGGAACTCCGTTGGGGTAAATTGACATCTGGGCCATTAACGGTGGAG 252
 Qy 61 GAAAGGCGCAAGCATGACGACGAGCTTCTTACGCTGAAAGCCGATGCGGATTTATCT 120
 Db 253 GAAAGGCTTAAACAAAGACGACGAGTTCATGCGGCTCAAGCCCAACGCTGTATATACA 312
 Qy 121 GGTGATCAAGCGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTGTAGCAAA 180
 Db 313 GGTGACCAAGCTAGGAAATTTTTCCTTACATCGGCTCTCCCGACCTGTGTGACAG 372
 Qy 181 ATATGGGCGCTAGCGGACATGATTAACGATGGAAGATGATCAAGTGAATTTTCCATA 240
 Db 373 ATATGGGCGCTGTGATGATGAAATGATGGAAGATGACCAACTGAGTTCTCCATA 432
 Qy 241 GCCATGAAGCTTATTAACCTGAAGTACAGGATTAAGTCTCCCTCCACACTTCCCT 300
 Db 433 GCTATGAATTAATTCACACTAAACTACAGGTTACCGGTGCTCTATTTCTCCCTTCC 492
 Qy 301 GTCATGAACAGCAACAGTGGCTAT-----TTCCAGTGCACGACATTTGTATAGA 354
 Db 493 AATATGTTGAAGCAACCGGTGTCATGCTGTGCTGTGTGGGGTGGGATGAGT 552
 Qy 355 GGGATGCTAGCATGCAACACTCAAGCTGTGCTGCTGTCGCAATGGCTTCAATTCCA 414
 Db 553 GGAATCGTGGGATCTCCCTAGCAGCTGAGCCCTGTGCAATGCAATCCATCCCA 612
 Qy 415 GTTGTGGAATGTCTCCACCTTGTATCTTGTGCTCTCCAGAGAGTGTGCTCCCTG 474
 Db 613 GTAGTAGGAATGTCTCTCTCTGTGTGTCTCTGTGCTCCCA-----CAGTTCCACCTTGG 666
 Qy 475 GCTAACGGGGGCTCTCCCGCTCAIACAGCCTGCGTGGTGGCATCTGCAGCCACA 534
 Db 667 TCTAACGGGGCTCTCGCAGTAACTACATCTCACTCTTGTGCTCACTCT---GCCACA 723
 Qy 535 TGGCCAAAGAGTCTTCTTACAGAGATCTGTGTCAGGGGTCACAATTAAACATAAGTTA 594
 Db 724 TTGCCAAAGAGTCTTCAATTGGCGGCTCAAGTGCAGGGTCCCAAGATTAACATAATT 783
 Qy 595 CAGAAGGACCAATCTTGTAGTGTGCGCAGCGCCCTCCAGAGAGAGATGGGCTGTGCT 654
 Db 784 CAGAAAGCTCAGCTCTTGTAGTGTGCAAGCCCCCACTTGTGTTGAATGGGCTGTGCA 843
 Qy 655 CAGTCATCAGGCTGAAATTAACAGGAGTTATTAACAGAGCAAGCAAAACTATAGTGA 714
 Db 844 TCGTCTTAAGATTGAAGTAAAGACAAATTTTCAACAGCAGAGCAAAACATGAGTGG 903
 Qy 715 CACTTAACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAG 774
 Db 904 AATTTAACAGGTCCTCAAGCAAGAACTATTCTTATGAGTCAAGTTTACCCCAAGCAG 963
 Qy 775 CTGGCTTAATATGGAATCTTCTTGAATGTGATCAAGATGAAAACTCACTGCAGAGAA 834
 Db 964 CTAGCTTAATAATATGGAATCTTCAACATTTGCAAGATGAAAACTCAAGCCAGCAAAA 1023
 Qy 835 TTATCTAGCTATGACCTAATTGATGTGCGATGTGCGAGCCAGCTGCGCGCTC 894
 Db 1024 TTATATTAAGCTATGACCTTAAATGATGTGCGCATGTGTGCGCAGCCACTTCTCTCAATC 1083
 Qy 895 CTGCTCCAGAAATACATCTCTCTCTCTTCAAGAAAGTGTGCTCGGAGTGGAGTGTCC 954

Db 1084 CTGCTCCAGAGTAAATTTCCCTCTTTTGAAGAGTGTGATCAGGAGTGGTTATCC 1143
 Qy 955 GTCAATAGCTCTTCTTGTGTGATCAGAGCGTGTGATGAGACCGGTGTCTAGAGATGAG 1014
 Db 1144 ATATGAGCTCTGTGTCTGTGATCAGCGCTTGTGCGAGGAAACAGAAAGAGAGACCA 1203
 Qy 1015 CAGCAGCAGAGAAAGAACTGCTGTGATCAATTTGAATTAAGAACCGGAGAACTTGCAG 1074
 Db 1204 CAAAACCGGACAAAGAACTCCAGTGAATTTGAAGCAAGAAAGAGAGAACTTTGAG 1263
 Qy 1075 CGAGCAGTGTGAGCTGTGAGAAAGCGCCGACCGCTCTTGTGAGCAGCAGCGCAAGAG 1134
 Db 1264 CGAGAAACCTAAGATTGAAGAAAGAGAGACAGGCACTTTTATGAGCAGCAGCGCAAGAG 1323
 Qy 1135 CAGAGGCGTGTGCTGTGAGACCGCGGAGAGAGAGAGAAAGCGGAGACGGCAG 1194
 Db 1324 CAGAGAGGTTGTGTCTGAGTGAAGAGGACAGAGAGAGAAAGAAACGAGACGGGAA 1383
 Qy 1195 GAGCAGAGAGCCAAAGCCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 1254
 Db 1384 GATCAAGAGCGGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1443
 Qy 1255 GAGCGGAGCGAG 1314
 Db 1444 GAGCGGAGAGCGGAG 1503
 Qy 1315 GAATGAAAGGAG 1374
 Db 1504 GAATGAAAGGAG 1563
 Qy 1375 CAG 1434
 Db 1564 CAAG 1623
 Qy 1435 TTGAGTTGAAGCTCTGAATGACAAAGAGATCAGTGAAGAGAACTTCAGAGATTC 1494
 Db 1624 TTGAGCTGGAAGCTCTGAATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 1683
 Qy 1495 AGTGTGAGCTGGAG 1554
 Db 1684 CGGTGTGCTCACTACTTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1743
 Qy 1555 AGAATGCTGAAATCACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
 Db 1744 AGAATGCAAGAGATCAACACTTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
 Qy 1615 AGACTTATTCAG 1674
 Db 1804 AAAATGATTCCTGAG 1863
 Qy 1675 TTGAGTGAAGAGCTGCTTCTTACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
 Db 1864 TTGCAAGAGAGATCTTGTCTTACACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1923
 Qy 1735 CAGCAGCTCCGGAG 1794
 Db 1924 CAGCAGCTCCGAGATCAAGTTAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1983
 Qy 1795 ATTGATGTTTCAACAACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
 Db 1984 ATCAGCTAATTATTAATCAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2043
 Qy 1855 CAGAAGAGAGAGTCCCTGTGAG 1914
 Db 2044 CAGAAAGAGAGAGCTTTGAACTGGAAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2103
 Qy 1915 CTGAGTTAG 1974
 Db 2104 AGCAGAGCTGATTAAGTGAAG 2163
 Qy 1975 TGCGTGAAGAGTGTGAG 2034
 Db 2164 TGCGAGAGTGGCTGAG 2217

OY	2035	AGACTGAAGGGGAAGACAGTGTCAAGAAAGAGGGCGGAAGAGGCCAAGCCGGA	2094
Db	2218	AAAGAAAAAGGAAAGAAATCCGCTCAGAAAGTCGAGGTGGAAAAA-----AAACCAAG	2271
OY	2095	ATGCAAGCAAGCAGAGTCGGCTTTTCATCCGCAATCAGAGCCAGCTACCTAGCCCAAC	2154
Db	2272	ATCCAGAGAAACCCAAACAAACCGTTCCACCAACTCCAGAGCCCTGGAAAACTAGAGGG	2331
OY	2155	CAGGACCCCTGGTCTACCAACAGAAAGGCCCGCTTACCAITTTCTGCACAGAGAGTGA	2214
Db	2332	CAGATTTCCCTGGATGATAATACGAGAAACCCCTCTMAAGAT---CAACACAGGGGATGTG	2388
OY	2215	AAAGGTATATATATCCGAGCGCTGATACCCTTTGAAATCCAGAAAGTCAAGATGATCAAC	2274
Db	2389	AAAGTGTGTATATACCGAGCACTTACCTTTTATGTCGGAAGTCAATGATAATCACT	2448
OY	2275	ATCCAGCCAGAGATATATAGTCATGCTGATGAAAGCCAGACTGAGAGCCAGATGGCTT	2334
Db	2449	ATTGAGCTCGRGAATATATCATGTGTGATGAAAGCCAAACYGAGAGACTGGGTGGCTG	2508
OY	2335	GGAGAGAGCTGAAAGGGGAAGCGGGATGTTCCCTGAAACTATATGCAAGAAATATCCA	2394
Db	2509	GGAGGTGAATCTGAAGGAAAAACGGGCTGGTTTCTGCAACTATGCAAGACGAATGCC	2568
OY	2395	GAAAAATGAGTTTCCACTCTCAAGCCAAACAGTGAACGATCTGACATCTGCCCTGCCCC	2454
Db	2569	GAGAGTGAATTTCCAAAGCAACATMAACAGAGAGAAATCAACAGCCAAACCACTGTCTC	2628
OY	2455	AAACTGTGCTCGCGTGAAGACCCCTGCTCTTTGCGCATGACCTCTTCTGAGCCCTCCA	2514
Db	2629	CATGTGGAC-----CTTCCCTGTGTCACTGTGAGTTTTCAMAAATACATCTACA	2679
OY	2515	ACCCCAACAACTGGGAGACTTCACTGTCCACGTGGCCACAGCTCAAGAGAAAGCA	2574
Db	2680	AATTCACAACTGGGAGACTTCACTGTCCACGTGGCCACAGCTCAAGAGAAAGCA	2739
OY	2575	GAAACGGAACCTGGGATACGTGGCGGCTCAGCCTTCTGTACCTGATGCTGGCC	2634
Db	2740	GAGAGTGAATCTGGGATACATGGGCACCTCAACATCTTCACTGTGCCAGTGTGTG	2799
OY	2635	CAGTTAGCGGAGAGATCAAGCCTTAAACCCACACACAGCACTGGCTCTCCCACTTCCC	2654
Db	2800	CAGCACAGACAGCGCTCTGTCTTTTACCCACACACTGTTACTGATCATGTGCCATCTCT	2859
OY	2695	GTCCTGGGCACAGGTGAAGAGGTGAAAGGGCTACAGAGCGCAAGCCCTGTATCCCTGAGA	2754
Db	2860	GTACTTGGTCAAGGGTGAAGAAAGTGAAGGCTTACAGGACAAAGCCTTGTACTTGGAGA	2919
OY	2755	GCCAAAAAAGCAACCACTTAAATTTTAACAAAAGTGACGTCATCACCGTCTGGAACAG	2814
Db	2920	GCAAAAGAAAGCAACCACTTAAATTTTAACAAAATGATGTATACAGGTCTTGAACAG	2979
OY	2815	CAAGACATGTGTGTGTGAGAAAGTTCAAGGTCAAGAGGGTGTGTTCCCAAGCTTAC	2874
Db	2980	CAGAGTATGTGTGTGTGAGAAAGTTCAAGGTCAAGAGGGCTGTGTTCCCAATCTCAT	3039
OY	2875	GTGAACATCTTTCAGGGCCGCTAAGAAATCCAACAGATCGATCTAGCCCTTACTGA	2934
Db	3040	GTAAAGCTTATATCCGGTCCAATAAGAAATCTTCAAGCATCGATCTTCAAGCTTTCAGAA	3099
OY	2935	AGTCCGTGTAGCTTAAAGAGAGTGGCTTCCCGGCGCGCAAGCAAGCAATTCOCGGAGA	2994
Db	3100	AGCCAGGAGATTTTAAAGAGTCTCTTACCAAGCATTTTAAACCTGCAATCCAAGAGAA	3159
OY	2995	GAGTTTATTCGATGATCACATACGAGAGTTCTGAGCAAGAGATTTTAACTTTCAACAA	3054
Db	3160	GAAATCATATCATGATACATATAGAGAGCAACAGCAAGAGACTTGACTTTTCAACAA	3219
OY	3055	GGGAGTGTGATTTGGTTTACCAAGAAATGATGATCTGTGAGAGGGAGCGGTGGAGAC	3114
Db	3220	GGAGTCTGATTTGGTGTATTTAAAGAAATGATGCACTGTGTGACTGTGACGGTGGAGAG	3279

QY	3115	AAGTCGGAGCTTCCCTTCTAACTAATGTAAGGCTTAAAGATTACAGAGGCTCTGAACT	3174
Db	3280	AAGATCGGGGTTTTCCCTTCCAACTAGCTAGGCCCAAAGACTCGAGGCGGCTGGAAGC	3339
QY	3175	GCTGGAAAACAGGAGATTAGAAAAAAACCTGAATTGGCCAGTTATTCTTCCAC	3234
Db	3340	GGAGAAAAGACAGMAAGCTTTGGGMAAACAACGAAATTGCCAGTTATTGCACTTAT	3399
QY	3235	GCTGCTACTGGTCCCGAAACAACCTACCTGGCTCTTGAGGACAGCATTTGATCCGAAA	3294
Db	3480	GCAGCACTGTCCTTGAAACAGCTAACACTGGCTCTGSCCACTTATCTTCATCCGAAA	3459
QY	3295	AAGAACCCAGGTGATGCTGGAGAGAGAACTGCMACTCGAGGAAAAAGCCGACATA	3354
Db	3460	AAGAAATCCAGAGGATGCTGGGAAGAGAGCTGCAGCAAGCTGGGAGAAAGACGCACATT	3519
QY	3355	GGAGGTTTTCCAGGAAATTATGTCAAACTTCTAAGCCCGGGAACAAACAAATACCCCA	3414
Db	3520	GGATGGTTTTCCAGCAAAATTATGTTAAAGCTTTGAGCCCTGTYACAAATTAAGACAGCCA	3579
QY	3415	ACTAGAGCTACCCAAAGACCGCAGTCAGCCAGCACTGTGCAAGTGATCGGATGTACAT	3474
Db	3580	ACKAAACCCWCCCAACCAACGATCTATTCCTCTCACTGACGCAAGTTATTCGGGATGTACAT	3639
QY	3475	TACACCGCCCAAGAACGATGACGAACTGACCTTCAGCAAAAGCCAGATCATCAAGTCTC	3534
Db	3640	TATATCGGCAAAATGACGAGAGTTGGCTTCAGTAAGAGGACGGGATTTAACGTCTG	3699
QY	3535	AACAAAGAGACCCCGACTGCTGTGGAAAGAGAAATCAGTGGGCAAGTTGGGCTTTCCCA	3594
Db	3700	AACAAAGAGATCCCGACTGCTGTGGAAAGAGAACTGAATGGCCATGTGTGCTGTTCCC	3759
QY	3595	TCCAAATTATGTAAAGCTGACCAAGACATAGAACCCAGCCAGCAAT	3640
Db	3760	TCCAACTACGTGAACCTGACCAAGACATGATCTTACGTAGCAAT	3805

RESULT	10
LOCUS	BD127168
DEFINITION	BD127168 2131 bp DNA linear PAT 18-SEP-2002
ACCESSION	BD127168
VERSION	BD127168.1 GI:23222113
KEYWORDS	JF 20020117375-A/2599.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2131) Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Naga,K., Kojima,S., Otsuki,T. and Koga,H. Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002017375-A 2599 22-JAN-2002; HELIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/2599 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO ISHII, PI YURI KAMAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA
TITLE	
JOURNAL	
COMMENT	
FEATURES	FT CDS Location/Qualifiers Location/Qualifiers (205)..(2130). I..2131
SOURCE	

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT      694 a      469 c      526 g      442 t
ORIGIN
Query Match      40.7%; Score 1483.8; DB 6; Length 2131;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

Qy 1 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCTGTAGTGTGGGCATTAAGTGGAG 60
Db 205 ATGGCTCAGTTTCCACACCTTTCGGTGTAGCCTGTAGTGTGGGCATTAAGTGGAG 264
Qy 61 GAAAGGCGCAGCATGACGACGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATCT 120
Db 265 GAAAGGCGCAGCATGACGACGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATCT 324
Qy 121 GGTGATCAAGGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTCTAGCACA 180
Db 325 GGTGATCAAGGAGGAACTTTTTCATCTGGGTTACCTCAGCCTGTCTAGCACA 384
Qy 181 ATATGGGCGCTAGCGGACATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 240
Db 385 ATATGGGCGCTAGCGGACATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 444
Qy 241 GGCATGAAGCTTATCAAACTGAAGCTACAGGATTCAGCTCCCTCCACATTTCCCT 300
Db 445 GGCATGAAGCTTATCAAACTGAAGCTACAGGATTCAGCTCCCTCCACATTTCCCT 504
Qy 301 GTCATGAAGCAGCAACAGTGGCTATTTCCAGTGCACAGCATTTGGTATGAGAGATT 360
Db 505 GTCATGAAGCAGCAACAGTGGCTATTTCCAGTGCACAGCATTTGGTATGAGAGATT 564
Qy 361 GCTAGCATGCCACCACTCAGACCTGTGCTCTGTGCCATGGGCTCCATTCAGATTGT 420
Db 565 GCTAGCATGCCACCACTCAGACCTGTGCTCTGTGCCATGGGCTCCATTCAGATTGT 624
Qy 421 GGAATGTCTCCACCTTATGATCTTCTGCTCCCTCAGACGACAGTGTCTCTGCTTAC 480
Db 625 GGAATGTCTCCACCTTATGATCTTCTGCTCCCTCAGACGACAGTGTCTCTGCTTAC 684
Qy 481 GGGGCTCCCTCCGTCATCAGGCTCTGCTGCTGGTGGCATCTGACGCCATAGGCA 540
Db 685 GGGGCTCCCTCCGTCATCAGGCTCTGCTGCTGGTGGCATCTGACGCCATAGGCA 744
Qy 541 AAGAGTCTTCTCTTACAGAGATCTGATCCAGGGTCACAACTTAACACTAAGTACAGAG 600
Db 745 AAGAGTCTTCTCTTACAGAGATCTGATCCAGGGTCACAACTTAACACTAAGTACAGAG 804
Qy 601 GCACATCATTTGATGTGGCCAGGCCCTCCAGACAGAGAAATGGGCTGTGCTCAGTCA 660
Db 805 GCACATCATTTGATGTGGCCAGGCCCTCCAGACAGAGAAATGGGCTGTGCTCAGTCA 864
Qy 661 TCAAGGCTGAATTAACAGGAGTTTCAACAGACGACGACAAACTATGATGAGGACATTA 720
Db 865 TCAAGGCTGAATTAACAGGAGTTTCAACAGACGACGACAAACTATGATGAGGAGATTA 924
Qy 721 ACAGGCTCCCAAGCAAGACTATTCATGCAATCAAGTTTACCCCAAGGCTCAGCTGCT 780
Db 925 ACAGGCTCCCAAGCAAGACTATTCATGCAATCAAGTTTACCCCAAGGCTCAGCTGCT 984
Qy 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCAGTCGAGAAATTTATC 840
Db 985 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCAGTCGAGAAATTTATC 1044
Qy 841 CTAGCTATGACCTTAATGATGTGCATGTGCGACGCGACCTGCGGCTCTGCTGCT 900
Db 1045 CTAGCTATGACCTTAATGATGTGCATGTGCGACGCGACCTGCGGCTCTGCTGCT 1104
Qy 901 CCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db 1105 CCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1164

```

```

Qy 961 AGCTCTTCTTCTGTGATCAGAGGCTGAGGAGCGCTGTCAGAGGATGACAGCAG 1020
Db 1165 AGCTCAACATCTGTAGGTCAGAGGCTACAGAGGAACTGTTTAGAAGATGAACACAA 1224
Qy 1021 C---CAGAGAAAGAACTGCTGTGATCTTGAATGAAGAGCGGAGAACTTTCAGCGA 1077
Db 1225 CAATTAGAAAAGAAATTAACCTGTAACTGTGAATGAAGAGCGGAGAACTTTGAACGT 1284
Qy 1078 GGCAGTGTGAGCTGTGAGAAAGCCGCCAAGCCTCTTTGAGACGACGCCGAAGAGCAG 1137
Db 1285 GGCAGCTGTGAGAACTGTGAGAAAGCAAGCAAGCTCTCTGTGAAACGACGCCAAGAGCAG 1344
Qy 1138 GAGGCGTTGGCTGAGCTGTGAGCGCGGAGAGGAGAAAGCGGAGCGGAGAG 1197
Db 1345 GAGGCGCTGGCCAGCTGTGAGCGGCGGAGAGGAGAGAGAGAGAGCTGAGCAGAG 1404
Qy 1198 CAGAGGCGCAAGCCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db 1405 CAAGAGGCGCAAAAGACAACTGTGAAGCACTGTGAAGAGCAAGCGGAGCTAGAA 1464
Qy 1258 CGGACGAGAGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAACGGGAA 1317
Db 1465 CGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
Qy 1218 CTGGAAGGCGAGCAACTTGAATGGGAATCGAATCGGAACCGGAGACAGGAATCTGAAATCAG 1377
Db 1525 CTGGAAGGCGAGCAACTTGAATGGGAATCGAATCGGAACCGGAGAGAGAGAGAG 1584
Qy 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 1585 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1644
Qy 1438 GAGTTAGAGCTGTGAATGACAAAAGATCAGCTAGAGAGAACTTTCAGGATATCAGG 1497
Db 1645 GAGTTAGAGCTGTGAATGACAAAAGATCAGCTAGAGAGAACTTTCAGGATATCAGG 1704
Qy 1498 TGTGAGCTGCGAACCCAGAGGAGAGAGAAATGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 1705 TGTGAGCTGCGAACCCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1764
Qy 1558 ATTGCTGAATCAACCACTTACAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 1765 ATTGCTGAATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1824
Qy 1618 CTATATTCAGAGAAACAGATCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 1825 CTATATTCAGAGAAACAGATCTCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1884
Qy 1678 CATAGAGCTGCGCTTCTTACCCCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
Db 1885 CATAGAGCTGCGCTTCTTACCCCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1944
Qy 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
Db 1945 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2004
Qy 1798 GATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Db 2005 GATGTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2064
Qy 1858 AAGCAGAGGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
Db 2065 AAGCAGAGGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124
Qy 1918 GAGTTAG 1924
Db 2125 GAATTAG 2131

```

```

RESULT 11
AK074554
LOCUS AK074554 2131 bp mRNA linear PRI 03-SEP-2002

```

```

DEFINITION Homo sapiens cDNA FLJ90073 fis, clone HEMBA1004110, highly similar
               to Homo sapiens interectin short form mRNA.
ACCESSION AK074554
VERSION AK074554.1 GI:22760070
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Bukaryova, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 Isogai, T., Ota, T., Nishikawa, T., Hayaashi, K., Otsuki, T.,
  Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,
  Kawai-Hio, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y.,
  Kojima, S., Nagahara, K., Masuko, Y., Ono, T., Okano, K., Yoshikawa, Y.,
  Aotsuma, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
  Nimomiya, K.
  NEDO human cDNA sequencing project
  Unpublished
  2 (bases 1 to 2131)
  Isogai, T. and Otsuki, T.
  Direct Submission
  Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
  Genomic Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
  (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
  NEDO human cDNA sequencing project supported by Ministry of
  Economy, Trade and Industry of Japan; cDNA full insert sequencing:
  Research Association for Biotechnology; cDNA library construction:
  Institute of Medical Science, University of Tokyo, Laboratory of
  Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
  sequencing and clone selection: Helix Research Institute (supported
  by Japan Key Technology Center etc.).
  Location/Qualifiers
  1..2131
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="HEMBA1004110"
  /issue_type="whole embryo, mainly head"
  /clone_lib="HEMBA1"
  /dev_stage="embryo, 10 weeks"
  /note="Cloning vector: pMB18SFL3"
  BASE COUNT 694 a 469 c 526 g 442 t
  ORIGIN
  Query Match 40.7%; Score 1483.8; DB 9; Length 2131;
  Best Local Similarity 86.0%; Pired. No. 0;
  Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;
  QY 1 ATGGCTCAGTTTCCACACCTTCGCTGGTAGCCTGATGTCGGCCATTAACCTGTGAG 60
  DB 205 ATGGCTCAGTTTCCACACCTTCGCTGGTAGCCTGATGTCGGCCATTAACCTGTGAG 264
  QY 61 GAAAGGGCAAGCATGACGAGCATGTTCTTACCTGAAGCCGATACGGGATTTATTA 120
  DB 265 GAAAGGGCAAGCATGATCAGCAGTTCATAGTTTAAAGCAATATCTGGATTCATTA 324
  QY 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTACTAGGCTGCTTGAACAA 180
  DB 265 GGTGATCAAGCGAGAACTTTTTCATCTGGGTACTAGGCTGCTTGAACAA 324
  QY 325 GGTGATCAAGCGAGAACTTTTTCATCTGGGTACTAGGCTGCTTGAACAA 384
  DB 181 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATCAAGTGAATTTTCATA 240
  QY 385 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATCAAGTGAATTTTCATA 444
  DB 241 GCCATGAAGCTTATCAAACTGAAGTACAGATATCACTCCCTCCACACTTCCCT 300
  QY 445 GCTATGAAGCTTATCAAACTGAAGTACAGATATCACTCCCTCCACACTTCCCT 504
  DB 301 GTCATGAAGCAGAACCAAGGCTATTTCCAGTGCACCGCATTGTTGATAGAGGATT 360
  QY 505 GTCATGAAGCAGAACCAAGGCTATTTCCAGTGCACCGCATTGTTGATAGAGGATT 564
  DB 361 GTCATGAAGCAGAACCAAGGCTATTTCCAGTGCACCGCATTGTTGATAGAGGATT 420

```

```

DB 565 GCCAGATGCCAACCGCTTACAGCTGTGCTCAGATGCCAATGGATCCATTCAGTTGTT 624
QY 421 GGAATGTCTCCACCTTATGATCTTGTGCTCCAGACGACGTGCTCCCTGGCTAAC 480
DB 625 GGAATGTCTCCACCTTATGATCTTGTGCTCCAGACGACGTGCTCCCTGGCTAAC 684
QY 481 GGGGCTCCCTGGCTTATAGGCTGCTGCTGGGTTGGGATCCCGACCCCATGGCCA 540
DB 685 GGGGCTCCCTGGCTTATAGGCTGCTGCTGGGTTGGGATCCCGACCCCATGGCCA 744
QY 541 AAGAGTCTTCCCTGAGCAGATCTGCTCAGGGTCACAATTAACCTAAGTTACAGAAG 600
DB 745 AAGAGTCTTCCCTGAGTATGATCTGCTCAGGGTCACAATTAACCTAAGTTACAGAAG 804
QY 601 GCACATATCTTGATGTGCGCAGCGCCCTCCAGACGAGAAATGGCTGTGCTCAGTCA 660
DB 805 GCACATATCTTGATGTGCGCAGCGCTCCAGACGAGAAATGGCTGTGCTCAGTCA 864
QY 661 TCAGGCTGAATTAACAGGAGTATTCACACCCACGACAAATTAATGATGACACTTA 720
DB 865 TCAGGCTGAATTAACAGGAGTATTCACACCCACGACAAATTAATGATGACACTTA 924
QY 721 ACAGGTCCCGCAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
DB 925 ACAGGTCCCGCAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 984
QY 781 TCATATGGAATCTTCTGACATTCATGATCAAGTGAAGAACTCACTGACAGAAATTTATC 840
DB 985 TCATATGGAATCTTCTGACATTCATGATCAAGTGAAGAACTCACTGACAGAAATTTATC 1044
QY 841 CTAGCTATGCACTTAATGATATGATGTCATGTCGTGTCAGCCACTGCGCCGCTGCT 900
DB 1045 CTAGCTATGCACTTAATGATATGATGTCATGTCGTGTCAGCCACTGCGCCGCTGCT 1104
QY 901 CCAGAAATACATCCCTCTCTCTTCTGAGAAAGTTCCTCCGAGAGTGGATGTCGCTCAT 960
DB 1105 CCAGAAATACATCCCTCTCTCTTCTGAGAAAGTTCCTCCGAGAGTGGATGTCGCTCAT 1164
QY 961 AGCTCTTCTCTGATGATGAGAGGCTGCTGAGAGAGCCGCTGCTCAAGATGAGACGAG 1020
DB 1165 AGCTCTTCTCTGATGATGAGAGGCTGCTGAGAGAGCCGCTGCTCAAGATGAGACGAG 1224
QY 1021 C---CAGAGAAAGAACTGCTGCTGATGATTTGAAGTAAAGACGGGAGAACTTCAGACGA 1077
DB 1225 CAATTAGAAAGAAAGAACTGCTGATGATTTGAAGTAAAGACGGGAGAACTTCAGACGA 1284
QY 1078 GGCAGTGTGAGCTGAGAGAGGCGCCAGAGGCTCTTGGAGCAGACGCGCAAGAGCGAG 1137
DB 1285 GGCAGTGTGAGCTGAGAGAGGCGCCAGAGGCTCTTGGAGCAGACGCGCAAGAGCGAG 1344
QY 1138 GAGCGGTTGCTCAGCTGAGAGCGCGCAGACGAGAGAGAAAGACGGGAGCGCAGAG 1197
DB 1345 GAGCGGTTGCTCAGCTGAGAGCGCGCAGACGAGAGAGAAAGACGGGAGCGCAGAG 1404
QY 1198 CAGAGAGCGCAAGCGGAGCTGAGAGCTGAGAGAGCGCTGAGAGAGCAGCGGAGCTGAG 1257
DB 1405 CAGAGAGCGCAAGCGGAGCTGAGAGAGCGGAGAGCGGAGAGCGGAGCTGAGAG 1464
QY 1258 CCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
DB 1465 CCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
QY 1318 CTGAAAGGACGACGACAACTTGAATGAGAGCGAAACGGAGACAGAACTCTGATACG 1377
DB 1525 CTGAAAGGACGACGACAACTTGAATGAGAGCGAAACGGAGACAGAACTCTGATACG 1584
QY 1378 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB 1585 AGGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1644
QY 1438 GAGTTAGAGAGCTGATATGACAAAGAGATCAGTGAAGAGAACTTCAGAGATATCAGG 1497

```


Db 1645 GAATTAGAGGCTCTAATATGATTAAGATCACTAGAGGAAAATTCAAGATATCAGA 1704
Qy 1498 TGTGACCTGCGACCCAGAGCAAGAAATTTGAGGACAGAACTCTAGAGGTAGA 1557
Db 1705 TGTGATTTGACACCCAAAGGCAAGAAATTTGAGGACAGAACTCTAGAGGTAGA 1764
Qy 1558 ATTCGTGAATACCCACTTACAGGAGAGTTCAGAGATCTCAGCAATCTTGGAGA 1617
Db 1765 ATTCGCAATACCCACTTACAGGAGAGTTCAGAGATCTCAGCAATCTTGGAGA 1824
Qy 1618 CTATTCAGAGAAACAGATATCTAGTACCACTTTAAACCAAGTCCAGAGAGTTG 1677
Db 1825 CTATTCAGAGAAACAGATATCTAGTACCACTTTAAACCAAGTCCAGAGAGTTG 1884
Qy 1678 CATGAGACTCGCTCTTACCTCAAGAGCTTTGAGAGCAAGAGAGTCCGCGAG 1737
Db 1885 CACAGAGATTCAGCTTTGTAAGAGCTTTGAGAGCAAGAGAGTCCGCGAG 1944
Qy 1738 CAGCTCCGAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
Db 1945 CACCTAG 2004
Qy 1798 GATGTTTCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Db 2005 GATGTTTCAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2064
Qy 1858 AAGCAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
Db 2065 AAGCAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124
Qy 1918 GAGTTAG 1924
Db 2125 GAATTAG 2131

RESULT 12
BD158570 2131 bp DNA linear PAT 17-JAN-2003
LOCUS BD158570 2131 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD158570.1 GI:27864328
VERSION JP 2002191363-A/13413.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Ota,T., Isegai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002191363-A 13413 09-JUL-2002;
HELIOS RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/13413
PD 09-JUL-2002
PI 28-JUL-2000 JP 2000280990
PI TOSHIO OTA, TAKAO ISEGA, TETSUO NISHIKAWA, KOJI HAYASHI, KAZUO
SAITO, JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,
PI KEIICHI NAGAI, TETSUJI OTSUKI
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ 10,
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N5/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1..2131
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 691 a 448 c 572 g 420 t

ORIGIN
Query Match 34.6%; Score 1259.2; DB 6; Length 2131;
Best Local Similarity 81.2%; Pred. No. 9.1e-275;
Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;
Qy 1 ATGAGCTAGTTTCCACACCTTTGAGTGTAGCTGTGATCTGAGGCAATATCTGTGAG 60
Db 347 ATGAGCTAGTTTCCACACCTTTTGGTGGAGCTGTGATCTGAGGCAATATCTGTGAG 406
Qy 61 GAAAGGCCAAGCATGACAGAGAGTTCCTAGCCTGAAGCCGATAGCGGATTTATTA 120
Db 407 GAAAGGCCAAGCATGATGATGACAGTTCATAGTTTAAAGCAATATCTGTGATTA 466
Qy 407 GAAAGGCCAAGCATGATGATGACAGTTCATAGTTTAAAGCAATATCTGTGATTA 466
Qy 121 GGTGATGAAGAGAGAGTCTTTTTCATCTGAGGTTTACTCTGAGCTGTTAGAGCA 180
Db 467 GGTGATGAAGAGAGAGTCTTTTTCATCTGAGGTTTACTCTGAGCTGTTAGAGCA 526
Qy 181 ATATGGCGCTAGCGGACATGAATTAACATGAGAGAGATGATCAAGTGAATTTCCATA 240
Db 527 ATATGGCGCTAGCGGACATGAATTAACATGAGAGAGATGATCAAGTGAATTTCCATA 586
Qy 241 GCATGAAGCTTATCAATCTGAGAGTCAAGATATGATGATGATGATGATGATGAT 300
Db 587 GCTATGAAGCTTATCAATCTGAGAGTCAAGATATGATGATGATGATGATGATGAT 646
Qy 301 GTCATGAAG 360
Db 647 GTCATGAAG 691
Qy 361 GCTAGCATGCGACCATCTACAGAGTGTGCTCTGTGCAATAGGAGCTCAATTCAGTTG 420
Db 692 ----- 691
Qy 421 GAAATGTCACACCTTAGATCTTCTGCTCCACAGAGAGTCTCCCTGCTGCTAC 480
Db 692 ----- 715
Qy 481 GGGGCTCCCTCCGATACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 716 GGGGCTCCCTCCGATACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
Qy 541 AAGAGTCTTCTTCCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 776 AAGAGTCTTCTTCCAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 835
Qy 601 GCACATATCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 836 GCACATATCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 895
Qy 661 TCAAGGCTGAATTCAGAGAGTTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 896 TCAAGGCTGAATTCAGAGAGTTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 955
Qy 721 ACAGGCTCCGAG 780
Db 956 ACAGGCTCCGAG 1015
Qy 781 TCAATATGAGAACTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 1016 TCAATATGAGAACTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1075
Qy 841 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 1076 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1135
Qy 901 CCAAGATATACCT 960
Db 1136 CCAAGATATACCT 1195
Qy 961 AGCTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 1196 AGCTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1255


```

QY 1021 C---CAGAGAGAAAGCTGCTGATGACATTTGAAGATPAGAAGCGGAGAACTTCGAGCGA 1077
Db 1256 CAATTGAAAAGAAATTACTGTATACCTTTGAAGATPAGAAGCGGAGAACTTTGAAGCT 1315
QY 1078 GGCAGGTGTGAGCTGAGAAAGCCCGCAAGCCCTCTTGAGAGACAGCGCAAGAGCAG 1137
Db 1316 GGCAGCTGTGAGCTGAGAAAGCGCAAGCTCTCTGAGAGAGCAGCGCAAGAGCAG 1375
QY 1138 GAGCGGTGTGAGCTGAGAGCGCGCGAGAGAGAGAAAGCGGAGCGCCCGAGAG 1197
Db 1376 GAGCGCTGTGAGCTGAGAGCGCGCGAGAGAGAGAGAGAGCGTGTGAGCGCCCGAGAG 1435
QY 1198 CAGAGAGCCCAAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db 1436 CAGAGAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495
QY 1258 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
Db 1496 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
QY 1318 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
Db 1556 CTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1615
QY 1378 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 1616 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
QY 1438 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1676 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
QY 1498 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 1736 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
QY 1558 ATTGCTGAAGATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 1796 ATTGCTGAAGATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
QY 1618 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 1856 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1915
QY 1678 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
Db 1916 CATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1975
QY 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
Db 1976 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2035
QY 1798 GATGTTTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Db 2036 GATGTTTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
QY 1858 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893
Db 2096 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

```

```

RESULT 13
AK027846 2131 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens cDNA FLJ14940 fis, clone PLACE1010942, highly similar
to Homo sapiens interseclin long isoform (ITSN) mRNA.
ACCESSION AK027846
VERSION AK027846.1 GI:14042823
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS
1 Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,
Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,
Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahori,K.,
Masuko,Y. and Kanehori,K.
TITLE
NEDO human cDNA sequencing project
JOURNAL
Unpublished
2 (bases 1 to 2131)
REFERENCE
AUTHORS
Isogai,T. and Otsuki,T.
TITLE
Direct Submision
JOURNAL
Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@ri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology; cDNA library construction;
5'- & 3'-end one pass sequencing and clone selection; Helix
Research Institute (supported by Japan Key Technology Center etc.)
and Department of Virology, Institute of Medical Science,
University of Tokyo.
FEATURES
Location/Qualifiers
source
1..2131
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1010942"
/tissue_type="Placenta"
/clone_1b="PLACE1"
/note="Cloning vector: pME18FL3"
BASE COUNT
691 a 448 c 572 g 420 t
ORIGIN
Query Match 34.6%; Score 1259.2; DB 9; Length 2131;
Best Local Similarity 81.2%; Pred. No. 9.1e-275;
Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;
QY 1 ATGGCTCAGTTTCCACACCTTTGGGTGATGCTGAGTGTGGCCATACTGTGGAG 60
Db 347 ATGGCTCAGTTTCCACACCTTTGGGTGATGCTGAGTGTGGCCATACTGTGGAG 406
QY 61 GAAAGGCCAAGACATGACGAGAGAGTTCCTTACCTGGAACCGATGCGGATTTATTA 120
Db 407 GAAAGGCCAAGACATGATGAGAGAGTTCCTTACCTGGAACCGATGCGGATTTATTA 466
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTTTACTCTAGCTGTCTTACAGCA 180
Db 467 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTTTACTCTAGCTGTCTTACAGCA 526
QY 181 ATATGGGCGCTTACGAGCATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 240
Db 527 ATATGGGCGCTTACGAGCATGATTAACATGAGAGATGATCAAGTGAATTTTCATA 586
QY 241 GCCATGAAGCTTATCAAACTGAAGCTACAGGATATACGCTCCCTCCACACTCCCT 300
Db 587 GCTATGAAGCTTATCAAACTGAAGCTACAGGATATACGCTCCCTCCACACTCCCT 646
QY 301 GTCATGAACACAGCAACAGTGTCTATTTTCAGTGAACAGCAATTTGTATAGAGGAT 360
Db 647 GTCATGAACACAGCAACAGTGTCTATTTTCAGTGAACAGCAATTTGTATAGAGGAT 691
QY 361 GCTAGAGTCCACCACTCAGAGCTGTGCTCTGTGCAATGGGCTCATTCAGTTGTT 420
Db 692 GCTAGAGTCCACCACTCAGAGCTGTGCTCTGTGCAATGGGCTCATTCAGTTGTT 691
QY 421 GGAATGTCTCCACCTTAGTATCTGTGCTCCAGAGAGAGTGTCTCCCTGGCTAAC 480
Db 692 GGAATGTCTCCACCTTAGTATCTGTGCTCCAGAGAGAGTGTCTCCCTGGCTAAC 715
QY 481 GGGGCTCTCCCGGTATACAGGCTGTGCTGTGGTGTGGCATCTGCAGCCATGGGCA 540
Db 716 GGGGCTCTCCCGGTATACAGGCTGTGCTGTGGTGTGGCATCTGCAGCCATGGGCA 775

```


D	324	GAAAGAGGGAAGCAGATCAGAGTTCATAGTTAAAGCCAAATATCTGATTCATTACT	383
Q	121	GATGATCAAGGAGGAACCTTTTTTCCATATCTGGGTACTCTCAGCTGTCTTAGCAAA	180
D	384	GATGATCAAGGTAGAAACCTTTTTTTTTCATCTGGGTACTCTCAGCTGTCTTAGCAAG	443
Q	181	ATATGGGGGTAGGCGGAGATGAATGAATGGAAGATGATGATGAAATTTTCCATA	240
D	444	ATATGGGACATGACGTGAATGAATGAATGGAAGATGATGAAATTTTCCATA	503
Q	241	GCCATGAAGCTTATCAAACTGAAGCTCAAGAGATATCAGCTCCCTCCACATTCCTCCCT	300
D	504	GCTATGAAGAACTTATCAAACTGAAGCTCAAGAGATATCAGCTACCTCTGCACTTCCCTCT	563
Q	301	GTCAATGAAGCAGCAACAGTGGCTATTTCCAGTGCACCAAGCATTTGTATAGAGGATT	360
D	564	GTCAATGAAGCAGCAACAGTGGCTATTTCCAGTGCACCAAGCATTTGTATAGAGGATT	623
Q	361	GCTAGCAATGCCAACCTCAGAGCTGTGCTCTGTGCGCAATAGGGCTCCATTCAGTTGTT	420
D	624	GCCAGCAATGCCAACCGCTTACAGCTGTGCTCCAGTGCACCAAGGATATCAATTCAGTTGTT	683
Q	421	GGAATGTCTCAACCTGTATGATCTTCTGCTCCAGAGAGAGTGCCTCCCTGGCTAC	480
D	684	GGAATGTCTCAACCTGTATGATCTTCTGCTCCAGAGAGAGTGCCTCCCTGGCTAC	743
Q	481	GGGGCTCTCCCTGATACAGCTTCGCTGGCTGGCTATTCCTGCAGCCACATGAGCCA	540
D	744	GGGGCTCTCCCTGATACAGCTTCGCTGGCTGGCTATTCCTGCAGCCACATGAGCCA	803
Q	541	AAGAGTCTTCTCTAGCAGATCTGCTCAGGCTCACAATTAACCTAAGTTACAGAG	600
D	804	AAGAGTCTTCTCTAGTATGATCTGCTCAGGCTCACAATTAACCTAAGTTACAGAG	863
Q	601	GCACATCATTCGATGATGCGCCAGCGCCCTCAGCAGAGATAGGGCTGTGCTCAGTCA	660
D	864	GCACATCATTCGATGATGCGCCAGCGCTCCACAGAGGAGAGTGGAGCTGTGCTCAGTCA	923
Q	661	TCAAGGCTGAATAACAGGCAAGTTATTAACAGCCAGCAAAACTATGAGTGAACCTTA	720
D	924	TCAAGGCTGAATAACAGGCAATTTATCAATGATGATGAACAACTATGAGTGAACCTTA	983
Q	721	ACAGGTCCCGCAGGCAAACTATTTCTATGCAATCAATGAATTTACCCAGGCTCAGCTGCT	780
D	984	ACAGGTCCCGCAGGCAAACTATTTCTATGCAATCAATGAATTTACCCAGGCTCAGCTGCT	1043
Q	781	TCAATATGGAATCTTTGCACTATGATCAAGATGGAATCTTACAGCAGAGAAATTTATC	840
D	1044	TCAATATGGAATCTTTGCACTATGATCAAGATGGAATCTTACAGCAGAGAAATTTATC	1103
Q	841	CTAGCTATGCACTTAATGATGTTGTCACATGCTGGTCAAGCACTGSCGCGCTCGTGCCT	900
D	1104	CTGGCAATGCACTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1163
Q	901	CCAGAAATCATCCCTCTCTCTTCCAGAAAGTTCGCTCCGCGAGTGGAGTGTCCGTCATA	960
D	1164	CCAGAAATCATCCCTCTCTCTTCCAGAAAGTTCGCTCCGCGAGTGGAGTGTCCGTCATA	1223
Q	961	AGCTCTTCTTCTGCGATCAGAGGCTGCTGAGAGCCCTGCTGATGAGAGATGAGCAGAG	1020
D	1224	AGCTCTTCTTCTGCGATCAGAGGCTGCTGAGAGCCCTGCTGATGAGAGATGAGCAGAG	1283
Q	1021	C---CAGAGAAAGAACTGCTGTGCAATTTGAAGATTAAGACGGGGAACCTTGCAGCGA	1077
D	1284	CAATTTAGAAAAGAAATTAACCTGTATACGTTTGAAGATTAAGACGGGGAACCTTGCAGCGA	1343
Q	1078	GAGAGTGTGAGCTGTGAGAGAGCGCGCAAGCGCTTTGAGCAGCAGCGCAAAAGCAG	1137
D	1344	GAGAGTGTGAGCTGTGAGAGAGCGCGCAAGCGCTTTGAGCAGCAGCGCAAAAGCAG	1403
Q	1138	GAGCGGTGTGCTCAGCTGAGAGGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG	1197
D	1404	GAGCGGTGTGCTCAGCTGAGAGGCGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG	1463

Oy		1198	CAGAGGCGCAACCCGCAGCTGGACTGTGAAGAAGCAGTGGAGAAGCAAGCGGAGCTGGAG	1257
Dd		1464	CAAGAGCGCAAAAGACAACCTGTGAACCTGAGAACGCACTGGAAAAGCACGGGAGCTTGAA	15230
Oy		1258	CGGAGCGCAAGAGAGAGAGAGAGAGAGAGAGATCGAAGGCCGCAAGCCGCAAAACGGGAA	1317
Dd		1524	CGGCAAGAGAGAGAGAGAGAGAGAGAGAAATTGAGAGAGCGAGAGGCTGCAAAACGGGAA	1583
Oy		1318	CTGGAAGAGGACGCAACACTTGAATGGGAACCGAACCGAGACAGAAATCTCTTAATCAG	1377
Dd		1584	CTTAAAGGCAACACCAACTTTGATGTGGAAACGGATGAAAGGCAAGAACTATAATCAA	1643
Dd		1644	AGAAACAAGAGACAGAGAGCATATGTTGTA CTG 1676	
RESULT 15				
AKO75290				
LOCUS				
DEFINITION			Homo sapiens cDNA FLJ30809 fis, clone Y79AA1000778, highly similar	PRI 03-SEP-2002
ACCESSION			AKO75290	
VERSION			AKO75290.1	GI:22761283
KEYWORDS			oligo capping; fis (full insert sequence).	
SOURCE			Homo sapiens (human)	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS			1. Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahari,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y., Aotsuka,S., Saeki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Nimomya,K.	
TITLE			NEDO human cDNA sequencing project	
JOURNAL			Unpublished	
REFERENCE			2 (bases 1 to 1676)	
AUTHORS			Isogai,T. and Otsuki,T.	
TITLE			Direct Submission	
JOURNAL			Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:genomicshri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT			NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert construction: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).	
FEATURES			Location/Qualifiers	
source			1..1676	
			/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
			/clone="Y79AA1000778"	
			/cell_line="y79"	
			/cell_type="retinoblastoma"	
			/clone_lib="Y79A1"	
			/note="cloning vector: PMR85FL3"	
BASE COUNT			485 a 381 c 458 g 352 t	
ORIGIN				
Query Match			30.3%; Score 1104.2; DB 9; Length 1676;	
Best Local Similarity			86.8%; Pred. No. 1,4e-239;	
Matches 1227; Conservative			0; Mismatches 183; Indels 3; Gaps 1;	
Oy		1	ATGGCTAGATTGCCAACCTTTTCGGTGTAGAGCTGTGATGTCTGGGCAATACGTGTGAG	60
Dd		264	ATGGCTAGATTGCCAACCTTTTCGGTGTAGAGCTGTGATGTCTGGGCAATACGTGTGAG	323

```

OY 61 GAAAGGCCAAGCATGACGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATCT 120
Db 324 GAAAGAGGAGGAGCATGATCAGCATGATTTCAATAGTTAAAGCCATATCTGGATTCATTTACT 383
OY 121 GGTGATCAAGGAGGAACTTTTTTTCATCTGGGTTACTCAGCTGTCTTAGCAACA 180
Db 384 GGTGATCAAGGAGGAACTTTTTTTCATCTGGGTTACTCAGCTGTCTTAGCAACA 443
OY 181 ATATGGGGGCTTAGGGAGCATGATTAAGATGGAAGATGAGTCAATGGAAATTTTCCATA 240
Db 444 ATATGGGCACTAGCTGACATGATATATATGGAAGAAATGAGTCAATGGAGTTTCCATA 503
OY 241 GCCATGAAGCTTATCAAACTGAAGCTAAGGATATCAGTCCCTCCACAATTCCTCCCT 300
Db 504 GGTATGAAGAACTTATCAAACTGAAGCTAAGGATATCAGTCCCTCCACAATTCCTCCCT 563
OY 301 GTCATGAAGCAAGCAACAGTGGCTATTTTCAGTGCACACAGCATTTGGTATAGAGGAGTT 360
Db 564 GTCATGAAGCAAGCAACAGTGGCTATTTTCAGTGCACACAGCATTTGGTATAGAGGATTC 623
OY 361 GGTACGATGCCCACTGACAGCTGTGCTCCTGTGGCCAAATGGGCTCCATTCATTTGTT 420
Db 624 GCGACGATGCCCACTGACAGCTGTGCTCCTGTGGCCAAATGGGATTCATTCAGTTGTT 683
OY 421 GGAATGTCTCACACCTTAGTATCTTCTGCTCCTCCAGCAGCAGTGCCTCCCTGGCTAAC 480
Db 684 GGAATGTCTCCCAACCTTAGTATCTTCTGCTCCTCCAGCAGCAGTGTGCTCCCTGGCTAAC 743
OY 481 GGGGCTCCTCCGTCATACAGCTGTGCTGCTGGGTTGGGCTCTGTCAGCCCATGSCCA 540
Db 744 GGGGCTCCTCCGTCATACAGCTGTGCTGCTGGGTTGGGCTCTGTCAGCCCATTTGCCA 803
OY 541 AAGAGTTCTCCTTCCAGAGATCTGTGCTCAGGGTCACAATTAACAATAAGTTATCAGAA 600
Db 804 AAGAGTTCTCCTTCCAGAGATCTGTGCTCAGGGTCACAATTAACAATAAGTTATCAGAA 863
OY 601 GCACATCATTTGATGTGCGCCAGCCCTCCAGCAGCAGAAATGGGCTGTGCTCAGTCA 660
Db 864 GCACATCATTTGATGTGCGCCAGGTGCCAGTCCACAGTGGGAGAGTGGGCTGTCTCAGTCA 923
OY 661 TCAGGCTGAATTAACAGGAGTTTATTAACAGCCACAGCAAAATATGATGAGTGAACATTA 720
Db 924 TCAGGCTGAATTAACAGGAGTTTATTAACAGCCACAGCAAAATATGATGAGTGAACATTA 983
OY 721 ACAGGTCCCGAGGCAAGAACTATTCATGCAATGCAATTAAGTTTACCCAGGCTCAGTGGCT 780
Db 984 ACAGGTCCCGAGGCAAGAACTATTCATGCAATGCAATTAAGTTTACCAAGGCTCAGTGGCT 1043
OY 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCAGTCAGAAAGATTTATC 840
Db 1044 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTTACAGCAGAGAAATTTATC 1103
OY 841 CTAGCTATGACCTTAATTGATGTTGCAATGTCTGCTCAGCCACTGCCGCTCTGCTCT 900
Db 1104 CTAGCTATGACCTTAATTGATGTTGCAATGTCTGCTCAGCCACTGCCGCTCTGCTCT 1163
OY 901 CCAGAAATCATCCCTCCCTCTCCCTTCCAGAAAGTTGCTCCGAGAGTGGAGATGCCCTCAT 960
Db 1164 CCAGAAATCATTCACCTCTCTTTTGAAGAGTTTCAATCTGSCATGATATCTGTGCATA 1223
OY 961 AGCTCTTCTCTGTGGATCAGAGGCTGCTGAGAGCCGCTCGTCAGAGGATGAGCAGCAG 1020
Db 1224 AGCTCAACATCTGTAGATCAGAGGCTACAGAGGAAACAGTTTGAAGATGAACACAA 1283
OY 1021 C---CAGAGAGAAACTGCTGTGACATTTGAAGATAGAGAGCGGAGAACTTTCAGCGA 1077
Db 1284 CAATTAGAAAAGAAATTAAGCTGTAAAGTTTGAAGATTAAGAGCGGAGAACTTGAACGT 1343
OY 1078 GGCAGTGTGAGAGTGAAGAGGCGCGCCAGAGGCTCTTGGAGCAGACAGCGCAAGAGCAG 1137
Db 1344 GGCAGACTGTGAGAGTGAAGAGGCGAGGCTCTTCTGGAACAGAGCGCAAGAGCAG 1403
OY 1138 GAGCGTTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGGAAAGAGCGGAGCGCCAGAG 1197

```

```

Db 1404 GAGCGCTGGCCAGCTGTGAGCGGGCGAGCAGAGAGAAAGAGCGCTGAGCGCCAGAG 1463
OY 1198 CAGAGGCCAAGCGGCACTGAGACTGAGAAAGCACTGAGAAAGCAGCGGAGCTGGAG 1257
Db 1464 CAAAGCGCAAAAGAACTGGAACCTGGAAGCACTGGAAGCAAGCGGAGCTGAGAA 1523
OY 1258 CGGCAAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCCAAACGGGAA 1317
Db 1524 CGGCAAGAGAGAGAGAGAGAGAAATTAAGAGCGCAAGCGCTGCAAAACGGGAA 1583
OY 1318 CTGAAAAGCAGCGCAACTTGAATGGGAACGAACCGGAGCAGGAACTTCTGAATCAG 1377
Db 1584 CTTGAAAAGCAGCGCAACTTGAATGGGAACGAATCGAAGCGCAAGAACTTAATCAA 1643
OY 1378 AGGAAACAAGAGCAGAGGAGGCAACCTGTGCTG 1410
Db 1644 AGGAAACAAGAGCAGAGGACATGATGTTACTG 1676

```

Search completed: December 4, 2003, 21:34:19
 Job time : 8623.82 secs

DR P-PSDB: AAY57444.
 XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 XX
 PS Claim 6, Page 38-40; 99pp; English.
 CC The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = Ee-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eap1s protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants), (II)-specific antibodies (Ab), sequences antisense to the (I)
 CC polynucleotide, agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eap1s complex, then binding dynamain to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;
 Query Match 99.5%; Score 5060; DB 21; Length 5082;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5082; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

DB 541 CCTCCACACTTCCCTGTCTATGAACAGCAACAGGGCTATTTCAGAGCAACAGCA 600
 QY 601 TTGGATATAGAGGAGATTGCTAGCATGCGACCACTCACAGCTGTGCTCTGTGCCAATG 660
 DB 601 TTGGATATAGAGGAGATTGCTAGCATGCGACCACTCACAGCTGTGCTCTGTGCCAATG 660
 QY 661 GGCTCCATTCAGTTGTTGGAATGCTCCACCCCTAGATCTGTGCTCCGAGCA 720
 DB 661 GGCTCCATTCAGTTGTTGGAATGCTCCACCCCTAGATCTGTGCTCCGAGCA 720
 QY 721 GTGCTCCCTGCTGCTTAACGGGGCTCTCCGCTCATACAGCTCTGCTGCTTTGCGAT 780
 DB 721 GTGCTCCCTGCTGCTTAACGGGGCTCTCCGCTCATACAGCTCTGCTGCTTTGCGAT 780
 QY 781 CCTGAGCCACATGCGCCAAAGATTCTTCTTCAGCAGATTGGTCCAGGCTCAATTA 840
 DB 781 CCTGAGCCACATGCGCCAAAGATTCTTCTTCAGCAGATTGGTCCAGGCTCAATTA 840
 QY 841 AACCTAAGTACGAAGGCAATCATCTGATGCGAGGGCCCTCCAGAGCA 900
 DB 841 AACCTAAGTACGAAGGCAATCATCTGATGCGAGGGCCCTCCAGAGCA 900
 QY 901 TGGGCTGTGCTCAGTCATCAAGGCTGAATATACAGGAGTTATCAACAGCCAGACAA 960
 DB 901 TGGGCTGTGCTCAGTCATCAAGGCTGAATATACAGGAGTTATCAACAGCCAGACAA 960
 QY 961 ACTATGATGGAACATTAAACAGTCCCGAGCAAACTATTCATGCAATCAAGTTTA 1020
 DB 961 ACTATGATGGAACATTAAACAGTCCCGAGCAAACTATTCATGCAATCAAGTTTA 1020
 QY 1021 CCCAGGCTGAGTGGCTTCAATGGAATCTTCTGACATTAAGATGAAGAACTC 1080
 DB 1021 CCCAGGCTGAGTGGCTTCAATGGAATCTTCTGACATTAAGATGAAGAACTC 1080
 QY 1081 ACTGACAGAAGATTATCTATGATGACCTTAATGATGTCATGCTGTCAGCA 1140
 DB 1081 ACTGACAGAAGATTATCTATGATGACCTTAATGATGTCATGCTGTCAGCA 1140
 QY 1141 CTGCGCGCGCTCTGCTCTCAAGATTAATCCCTCTCTTCTCAGAAAGTTCGCTCGGC 1200
 DB 1141 CTGCGCGCGCTCTGCTCTCAAGATTAATCCCTCTCTTCTCAGAAAGTTCGCTCGGC 1200
 QY 1201 AGTGGAGTCTCCGCTATGAGCTCTTCTGATGAGAGGCGCTGAGAGAGCGCTG 1260
 DB 1201 AGTGGAGTCTCCGCTATGAGCTCTTCTGATGAGAGGCGCTGAGAGAGCGCTG 1260
 QY 1261 TCAGAGATGAGCAGAGCCAGAGAAACTGCTGTGACATTTGAAGATPAAGAGCG 1320
 DB 1261 TCAGAGATGAGCAGAGCCAGAGAAACTGCTGTGACATTTGAAGATPAAGAGCG 1320
 QY 1321 GAGAACTTTCAGAGGAGGAGTGTGAGTGGAGAAAGCGCCCAAGGCTTTTGAAGCAG 1380
 DB 1321 GAGAACTTTCAGAGGAGGAGTGTGAGTGGAGAAAGCGCCCAAGGCTTTTGAAGCAG 1380
 QY 1381 CAGGCAAAAGCAGAGCGGTTGGCTCAGCTGAGAGCGCGCAGCAGAGAGAAAGAG 1440
 DB 1381 CAGGCAAAAGCAGAGCGGTTGGCTCAGCTGAGAGCGCGCAGCAGAGAGAAAGAG 1440
 QY 1441 CCGAGAGCCCAAGAGCAGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 DB 1441 CCGAGAGCCCAAGAGCAGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1501 CAGCGGAGGCTGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 DB 1501 CAGCGGAGGCTGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1561 GCGGCAAAACGGGAACTGGAAAGCGACGCAACTTGAATGGGAAACGGAGAGAG 1620
 DB 1561 GCGGCAAAACGGGAACTGGAAAGCGACGCAACTTGAATGGGAAACGGAGAGAG 1620
 QY 1621 GAACTCTGATTCAGAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680

Db 1621 GAACCTCTGAATCAGAGGAAACAAGGACAGAGGCGACCGTGCTCTGAGGCAAGGAGG 1680
Qy 1681 AAGACTCTGAGATTGAGTTAGAGCTGGAATGACAAAAGATCAGCTAGAGGAAA 1740
Db 1681 AAGACTCTGAGATTGAGTTAGAGCTGGAATGACAAAAGATCAGCTAGAGGAAA 1740
Qy 1741 CTTGAGGATATCAGGCTGCTGAGCTGGCAACCCAGAGGCAAGAAATTGAGACGAAACAG 1800
Db 1741 CTTGAGGATATCAGGCTGCTGAGCTGGCAACCCAGAGGCAAGAAATTGAGACGAAACAG 1800
Qy 1801 TTTAGAGACTTAAGAAATTCCTGAATCACTTACAGCAAGAGTTGAGAAATCTCAG 1860
Db 1801 TTTAGAGACTTAAGAAATTCCTGAATCACTTACAGCAAGAGTTGAGAAATCTCAG 1860
Qy 1861 CAATGCTTGAAGAATTATTCAGAGAAACAGATCTAGTGAACAGTTAAACAAGTC 1920
Db 1861 CAATGCTTGAAGAATTATTCAGAGAAACAGATCTAGTGAACAGTTAAACAAGTC 1920
Qy 1921 CAGCAGAACAGTTTGTGATAGAGACTGCTTCTTACCTCAAAAAGACCTTGGAGCAAG 1980
Db 1921 CAGCAGAACAGTTTGTGATAGAGACTGCTTCTTACCTCAAAAAGACCTTGGAGCAAG 1980
Qy 1981 GAGCTGCTCCGCGCAGAGCTCCGGGAGAGCTGAGCAGAGTGAAGAGAGAGACAGTCA 2040
Db 1981 GAGCTGCTCCGCGCAGAGCTCCGGGAGAGCTGAGCAGAGTGAAGAGAGAGACAGTCA 2040
Qy 2041 AAGCTGAGGAGATTGATGTTTTCACAAACAGCTGAAGAGAACTGAGAAATACATAC 2100
Db 2041 AAGCTGAGGAGATTGATGTTTTCACAAACAGCTGAAGAGAACTGAGAAATACATAC 2100
Qy 2101 AAACAGCAACTCCAGAGAGAGAGTCCCTGAGAGCAGCGACTGAAGCAGAAAGAGCAG 2160
Db 2101 AAACAGCAACTCCAGAGAGAGAGTCCCTGAGAGCAGCGACTGAAGCAGAAAGAGCAG 2160
Qy 2161 GAGAGAGAGAGCTTGAAGTTAGAGAGCAAGAAAGAAAGAGCTTCAAGAGCAGTTCA 2220
Db 2161 GAGAGAGAGAGCTTGAAGTTAGAGAGCAAGAAAGAAAGAGCTTCAAGAGCAGTTCA 2220
Qy 2221 AAGGCAAGCAATGCTGAGAGATGTCAGCAGAGAGAGCAGCAGCCCCCGGAAACCC 2280
Db 2221 AAGGCAAGCAATGCTGAGAGATGTCAGCAGAGAGAGCAGCAGCCCCCGGAAACCC 2280
Qy 2281 CACGAGAGAGAGAGCTGAAGAGAGAGAGCAGTGTCAAGAGAAAGAGAGAGAGAGAG 2340
Db 2281 CACGAGAGAGAGAGCTGAAGAGAGAGAGCAGTGTCAAGAGAAAGAGAGAGAGAGAG 2340
Qy 2341 GCGAAGCCGGAATGCAAGAGAGAGAGAGTCCGCTTTTCATCCGATCAGAGCCAGCT 2400
Db 2341 GCGAAGCCGGAATGCAAGAGAGAGAGAGTCCGCTTTTCATCCGATCAGAGCCAGCT 2400
Qy 2401 AAGCTGGCCACCCAGAGCAGCTGCTACCAAGAGAAAGGCGGCTTACATTTCTGCA 2460
Db 2401 AAGCTGGCCACCCAGAGCAGCTGCTACCAAGAGAAAGGCGGCTTACATTTCTGCA 2460
Qy 2461 CAGGAGAGTGAAGAGTGAATTAACGAGGCTGTACCCCTTTGAATCCAGAGTCAAC 2520
Db 2461 CAGGAGAGTGAAGAGTGAATTAACGAGGCTGTACCCCTTTGAATCCAGAGTCAAC 2520
Qy 2521 GATGAGATCAACATCAGCAGAGAGATATAGTCATGTGATGAAAGCCAGACTGAGAG 2580
Db 2521 GATGAGATCAACATCAGCAGAGAGATATAGTCATGTGATGAAAGCCAGACTGAGAG 2580
Qy 2581 CAGAGATGCTTGGAGAGAGAGCTGAAGAGAGAGAGAGAGTGGTCCCTTCAAACTATGCA 2640
Db 2581 CAGAGATGCTTGGAGAGAGAGAGCTGAAGAGAGAGAGAGAGTGGTCCCTTCAAACTATGCA 2640
Qy 2641 GAAAGATTCAGAGAAATGAGGTTCCCACTCAGCCCAACAGTGAACGATCAGCATCT 2700
Db 2641 GAAAGATTCAGAGAAATGAGGTTCCCACTCAGCCCAACAGTGAACGATCAGCATCT 2700
Qy 2701 GCGCTGCCCCCAACTGGCTCTGCGTGAAGCCCTGCTCTTTGCAAGTGAACCTTCT 2760
Db 2701 GCGCTGCCCCCAACTGGCTCTGCGTGAAGCCCTGCTCTTTGCAAGTGAACCTTCT 2760

Qy 2761 GAGCCCTCCAGAACCCCAACAACTGGGAGAGCTTCACTTCCAGCTGGGCCAGAGCTCA 2820
Db 2761 GAGCCCTCCAGAACCCCAACAACTGGGAGAGCTTCACTTCCAGCTGGGCCAGAGCTCA 2820
Qy 2821 AACGAGAGCCAGAAACCGAGCAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCGTA 2880
Db 2821 AACGAGAGCCAGAAACCGAGCAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCGTA 2880
Qy 2881 CTTAGTCTGGCAGATTACGGCAGATCAGCTTACCCAGCCACAGCAGCTGCTCC 2940
Db 2881 CTTAGTCTGGCAGATTACGGCAGATCAGCTTACCCAGCCACAGCAGCTGCTCC 2940
Qy 2941 TCCCATGCTCCGCTGCTGGGCGCAGGAGTGAAGAGAGGCTTCAAGCCGCAAGCCCTG 3000
Db 2941 TCCCATGCTCCGCTGCTGGGCGCAGGAGTGAAGAGAGGCTTCAAGCCGCAAGCCCTG 3000
Qy 3001 TATCCCTGAGAGCCCAAAAAGACAAACCACTTAATTTTAAACAAAAGTGAAGTCAAC 3060
Db 3001 TATCCCTGAGAGCCCAAAAAGACAAACCACTTAATTTTAAACAAAAGTGAAGTCAAC 3060
Qy 3061 GTTCTGAAACAGCAAGACATGTGTGTTGAGAGATTCAAGTCAAGAGGTTGCTTC 3120
Db 3061 GTTCTGAAACAGCAAGACATGTGTGTTGAGAGATTCAAGTCAAGAGGTTGCTTC 3120
Qy 3121 CCCAGTCTTACGTAAACTCATTTCAAGGCCCCCAAGAAATCCACAAAGCATGATACT 3180
Db 3121 CCCAGTCTTACGTAAACTCATTTCAAGGCCCCCAAGAAATCCACAAAGCATGATACT 3180
Qy 3181 GCGCCTTACTGAAGTCTGCTAGTCTAAGAGAGAGTGTCCCGCGCCCAAGCCAGCC 3240
Db 3181 GCGCCTTACTGAAGTCTGCTAGTCTAAGAGAGAGTGTCCCGCGCCCAAGCCAGCC 3240
Qy 3241 ATTCCCGAGAGAGATTATTCGATGACATACATACATACATACATACATACATACAT 3300
Db 3241 ATTCCCGAGAGAGATTATTCGATGACATACATACATACATACATACATACATACAT 3300
Qy 3301 ACCCTTCAAGCAAGGAGATGATGTTGGTTCACAAAGAAAGATGATGATGATGATGAT 3360
Db 3301 ACCCTTCAAGCAAGGAGATGATGTTGGTTCACAAAGAAAGATGATGATGATGATGAT 3360
Qy 3361 ACGGTGGGCGCAAGATCCGAGAGTCTTCCCTTCTAATCTATGAGAGCTTAAAGATTCA 3420
Db 3361 ACGGTGGGCGCAAGATCCGAGAGTCTTCCCTTCTAATCTATGAGAGCTTAAAGATTCA 3420
Qy 3421 GCGCTGGAACCTGCTGGGAAAAACAGGAGTTTGAAGAAAAAACTGAAATTCAGAGTT 3480
Db 3421 GCGCTGGAACCTGCTGGGAAAAACAGGAGTTTGAAGAAAAAACTGAAATTCAGAGTT 3480
Qy 3481 ATTGCTTCTTACGCTGCTTACGCTGCTTACGCTGCTTACGCTGCTTACGCTGCTT 3540
Db 3481 ATTGCTTCTTACGCTGCTTACGCTGCTTACGCTGCTTACGCTGCTTACGCTGCTT 3540
Qy 3541 CTGATCCGGAAGAAAAAAGAACCCAGGTGATGAGTGAAGAGAGAACTCAAGCTGAGAGAA 3600
Db 3541 CTGATCCGGAAGAAAAAAGAACCCAGGTGATGAGTGAAGAGAGAACTCAAGCTGAGAGAA 3600
Qy 3601 AAGGCGCAGATAGGATGTTTTCAGCAAAATATGCAAACTTTCAAGCCCGGAAACAGC 3660
Db 3601 AAGGCGCAGATAGGATGTTTTCAGCAAAATATGCAAACTTTCAAGCCCGGAAACAGC 3660
Qy 3661 AAAATCACCACCACTGAGACTACCCAAAGCCGAGTGAAGCAGAGTGTCCAGAGTGTATC 3720
Db 3661 AAAATCACCACCACTGAGACTACCCAAAGCCGAGTGAAGCAGAGTGTCCAGAGTGTATC 3720
Qy 3721 GGGATGTAAGATTAACAGCCAGAGATGATGATGATGATGATGATGATGATGATGATG 3780
Db 3721 GGGATGTAAGATTAACAGCCAGAGATGATGATGATGATGATGATGATGATGATGATG 3780
Qy 3781 ATCAAGCTTCTCAACAGAGAGAGCCCGGATGATGATGATGATGATGATGATGATGAT 3840
Db 3781 ATCAAGCTTCTCAACAGAGAGAGCCCGGATGATGATGATGATGATGATGATGATGAT 3840


```

Oy 3841 GGGCTCTCCATCCATTTATGTAAAGCTGACACAGACATGAGCCCGACAGCAATGA 3900
Db 3840 GGGCTCTTCCATCCATTTATGTAAAGCTGACACAGACATGAGCCCGACAGCAATGA 3899
Oy 3901 ATCATATGTGTGTCATCCCGCTTCAGGCTTGAAAGTCTCAAGAGAGACCCACTATCCCA 3960
Db 3900 ATCATATGTGTGTCATCCCGCTTCAGGCTTGAAAGTCTCAAGAGAGACCCACTATCCCA 3959
Oy 3961 TATCATGCGCCAGAGAGAGATGATGGAGATGACAGCTTATCATGTGACTTTCAGACATGAT 4020
Db 3960 TATCATGCGCCAGAGAGAGATGATGGAGATGACAGCTTATCATGTGACTTTCAGACATGAT 4019
Oy 4021 CACCTACTGCTTGTGATGAGAGAACTCACTGACAGACAGTTTACTATTGACCTTA 4080
Db 4020 CACCTACTGCTTGTGATGAGAGAACTCACTGACAGACAGTTTACTATTGACCTTA 4079
Oy 4081 GTTGATGTGATTCGAAATGTCTGACTCACTGCTGACAGAGCAGAGAAATTCAGAAC 4140
Db 4080 GTTGATGTGATTCGAAATGTCTGACTCACTGCTGACAGAGCAGAGAAATTCAGAAC 4139
Oy 4141 TGCAAGGGTGTGGGTCTTTTGGGGCTTTCCTAGTCACTCAAGCTGACCGGCCGCC 4200
Db 4140 TGCAAGGGTGTGGGTCTTTTGGGGCTTTCCTAGTCACTCAAGCTGACCGGCCGCC 4199
Oy 4201 TTCACACGGGCGCTTTCATAGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 4260
Db 4200 TTCACACGGGCGCTTTCATAGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 4259
Oy 4261 ATAAATTCGATCAATTAATCTCTTTCCTTATTTTGTGTTTAAACCAACCCACTATC 4320
Db 4260 ATAAATTCGATCAATTAATCTCTTTCCTTATTTTGTGTTTAAACCAACCCACTATC 4319
Oy 4321 AAGAGTGCCTGTCTGCGGAGATTAATAATGTCTCCGGGCGTACCGTTAACTGAGAGC 4380
Db 4320 AAGAGTGCCTGTCTGCGGAGATTAATAATGTCTCCGGGCGTACCGTTAACTGAGAGC 4379
Oy 4381 TTGCTGTACTTTCGCTTTCCTGAGTGTCCCAACACATTTGTGTAGTTGGGGCTTGT 4440
Db 4380 TTGCTGTACTTTCGCTTTCCTGAGTGTCCCAACACATTTGTGTAGTTGGGGCTTGT 4439
Oy 4441 CCTGCGCTGAGACACAGAGAGATGGGTGTAACCTGTTTGAATAATGTGTAGTACGTG 4500
Db 4440 CCTGCGCTGAGACACAGAGAGATGGGTGTAACCTGTTTGAATAATGTGTAGTACGTG 4499
Oy 4501 AGCGTACTATGAGAGGGGTTATGCTGTCTGTGACCATCAGCTTACTCTGGGCGATG 4560
Db 4500 AGCGTACTATGAGAGGGGTTATGCTGTCTGTGACCATCAGCTTACTCTGGGCGATG 4559
Oy 4561 TACCATCTGTACCGAAGAGTAGTCTTCTCTCATGAGCTTAAACCCAGCACCGTGTACGT 4620
Db 4560 TACCATCTGTACCGAAGAGTAGTCTTCTCTCATGAGCTTAAACCCAGCACCGTGTACGT 4619
Oy 4621 GCTCTCATCTACTGCAATTCATTTTACTTTGACAGCTGACCTTGTAGCCACTGAGAGC 4680
Db 4620 GCTCTCATCTACTGCAATTCATTTTACTTTGACAGCTGACCTTGTAGCCACTGAGAGC 4678
Oy 4681 ACCCATGTTTCCGTTTGTGCTGACATGTAACCTAGTTGGCCCGGTGTTTGTATTTT 4740
Db 4679 ACCCATGTTTCCGTTTGTGCTGACATGTAACCTAGTTGGCCCGGTGTTTGTATTTT 4738
Oy 4741 TCAATCTGGCATGTCTTCAACACATTAAGTAGAGAGCCCACTGCCAGCGGTACG 4800
Db 4739 TCAATCTGGCATGTCTTCAACACATTAAGTAGAGAGCCCACTGCCAGCGGTACG 4798
Oy 4801 ATCATCACTACCCACCGTCTTATAGTCTGTGTAACGTGAAGTTTATTCAGTTGCTTTTAT 4860
Db 4799 ATCATCACTACCCACCGTCTTATAGTCTGTGTAACGTGAAGTTTATTCAGTTGCTTTTAT 4858
Oy 4861 GGAATATCTTGAACAAGTAATCTTCTTGACAAAGAAAGATGTATGAAGTCTCCCTGAA 4920
Db 4859 GGAATATCTTGAACAAGTAATCTTCTTGACAAAGAAAGATGTATGAAGTCTCCCTGAA 4918
Oy 4921 TTAATTTCCAGTGTGTTACATTTTAACTAGACTGTGGGGGTCTACAGATTAATG 4980

```

```

Db 4919 TTAATTTCCAGTGTGTTACATTTTAACTAGACTGTGGGGGTGTGACAGATTAATG 4978
Oy 4981 AATGCGGCTCTGCTGCTGCTGTGTGTGTTAACTTGTCTGTAGTGAAGCCGTGTCTT 5040
Db 4979 AATGCGGCTCTGCTGCTGCTGTGTGTGTTAACTTGTCTGTAGTGAAGCCGTGTCTT 5038
Oy 5041 TAGATATTAGTGAAGTCCGGAAGAGAAATTCGATCAAGCTT 5084
Db 5039 TAGATATTAGTGAAGTCCGGAAGAGAAATTCGATCAAGCTT 5082

RESULT 2
AAZ39024
ID AAZ39024 standard; cDNA; 5738 BP.
XX
AC AAZ39024;
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse EseiL cDNA sequence.
XX
KW Mouse; murine; Esei; Esei2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
OS Mus sp.
XX
PN M09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Wang W, Sengar A;
XX
PI WPI: 2000-052802/04.
XX
DR P-PSDB; AAY57449.
XX
PT New nucleic acid encoding Esei and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection.
XX
PS Claim 6; Page 56-59; 99pp; English.
XX
CC The present invention specifically describes mammalian Esei and 2
CC proteins (I) and their splice variants (Esei = EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-
CC mediated endocytosis (as a complex with Eps15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its antagonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Esei genes or antagonists of an Esei binding partner are
CC used to treat diseases associated with undesirable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Esei is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Esei mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Esei-Eps15 complex, then binding
CC dynam to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse EseiL cDNA sequence.
XX

```


SQ Sequence 5738 BP; 1556 A; 1456 C; 1613 G; 1113 T; 0 other;

Query Match 76.4%; Score 3886.4; DB 21; Length 5738;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 3898; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 1 CGGCAAGAGAGAGTGGAGCGGCGC - GGGAGGCGCGCAGCTGGTCTCCGTAGTAC 59
DB 1 CGGCAAGAGAGAGTGGAGCGGCGCGGGAGGGCGCGCAGCTGGTCTCCGTAGTAC 60
QY 60 GCGCGCTCGCAAGAGAGCATCCCGAGCGGGCTCCGGGACGCGCCGGAGGCGAGCGCG 119
DB 61 GCGCGCTCGCAAGAGAGCATCCCGAGCGGGCTCCGGGACGCGCGGGAGGCGAGCGCG 120
QY 120 GCGGCGGGGAGTGGTGGCGGCTGCGGACCTCGGCTTCCTCGCGCGCGCGGCGGCTG 179
DB 121 GCGGCGGGGAGTGGTGGCGGCTGCGGACCTCGGCTTCCTCGCGCGCGCGGCGGCTG 180
QY 180 CACTGATTTGTGTGAGGGGCGGCGCGCGACCCCGCGGAGATGAGCGGTGCATCAGA 239
DB 181 CACTGATTTGTGTGAGGGGCGGCGCGCGACCCCGCGGAGATGAGCGGTGCATCAGA 240
QY 240 AGGTAAAGTAATAGAACCATGCTCAGTTTCCACACCTTTCCGTTGAGCTGAGTGT 299
DB 241 AGGTAAAGTAATAGAACCATGCTCAGTTTCCACACCTTTCCGTTGAGCTGAGTGT 300
QY 300 CTGGGCGCAATCTGTGAGGAGAAAGGCGCAAGCATGACCGAGTTCCTTAGCTGAAGC 359
DB 301 CTGGGCGCAATCTGTGAGGAGAAAGGCGCAAGCATGACCGAGTTCCTTAGCTGAAGC 360
QY 360 GATAGCGGAGTTTATTACTGTGATCAAGCGAGAACTTTTTCCTTCCATCTGGGTACC 419
DB 361 GATAGCGGAGTTTATTACTGTGATCAAGCGAGAACTTTTTCCTTCCATCTGGGTACC 420
QY 420 TCAGCTCTGCTTAGACAACAAATATGCGCGCTAGCGGACATGAATTAAGTAGAGTGA 479
DB 421 TCAGCTCTGCTTAGACAACAAATATGCGCGCTAGCGGACATGAATTAAGTAGAGTGA 480
QY 480 TCAATGTGAATTTTTCATAGCCATGAAAGTTATCAAACTGAAGCTTCAAGATATCAGT 539
DB 481 TCAATGTGAATTTTTCATAGCCATGAAAGTTATCAAACTGAAGCTTCAAGATATCAGT 540
QY 540 CCCCTCCACACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACACAC 599
DB 541 CCCCTCCACACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACACAC 600
QY 600 ATTGGTATAGAGGAGATTGCTAGCAATGACCACTCAGAGCTGTGCTCTGTGCAAT 659
DB 601 ATTGGTATAGAGGAGATTGCTAGCAATGACCACTCAGAGCTGTGCTCTGTGCAAT 660
QY 660 GGGCTTCATTTCAAGTTTGGAAATGCTCCACCCCTTAGATCTTCTGCTCCACAGCAGC 719
DB 661 GGGCTTCATTTCAAGTTTGGAAATGCTCCACCCCTTAGATCTTCTGCTCCACAGCAGC 720
QY 720 AGTGCCTCCCTGGCTAACGGGGCTCTCCCGTCAATACAGCCCTGCGGCTTTGGCGA 779
DB 721 AGTGCCTCCCTGGCTAACGGGGCTCTCTCCGTCAATACAGCCCTGCGGCTTTGGCGA 780
QY 780 TCCTGAGGACATGCGCAAGAGTCTTCTTTCAGCAGATCTGTGTCAGGGTCAAT 839
DB 781 TCCTGAGGACATGCGCAAGAGTCTTCTTTCAGCAGATCTGTGTCAGGGTCAAT 840
QY 840 AAACTAAGTTACAGAGGCAATCATTTGATGTCGACAGCGCCCTCCAGCAGCAGA 899
DB 841 AAACTAAGTTACAGAGGCAATCATTTGATGTCGACAGCGCCCTCCAGCAGCAGA 900
QY 900 ATGGGCTGGCTCAGTCAAGGCTGAATATACAGGAGTATTCAACAGGCAAGCA 959
DB 901 ATGGGCTGGCTCAGTCAAGGCTGAATATACAGGAGTATTCAACAGGCAAGCA 960
QY 960 AACTATGAGTGAACCTTAACAGGTCCCGAGGCAAGAACTATTTCTATGCAATCAAGTTT 1019
DB 961 AACTATGAGTGAACCTTAACAGGTCCCGAGGCAAGAACTATTTCTATGCAATCAAGTTT 1020
```

```
QY 1020 ACCCGAGCTCAGCTGGCTTCAATATGGAATCTTTCGACATGATGCAAGATGAAAACT 1079
DB 1021 ACCCGAGCTCAGCTGGCTTCAATATGGAATCTTTCGACATGATGCAAGATGAAAACT 1080
QY 1080 CACTCAGAGAAATTTATCTAGCTATGACCACTTAATGATGTCATGTCGTGACGC 1139
DB 1081 CACTCAGAGAAATTTATCTAGCTATGACCACTTAATGATGTCATGTCGTGACGC 1140
QY 1140 ACTGCCGCCGCTCTGCTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1199
DB 1141 ACTGCCGCCGCTCTGCTCCAGAAATACATCCCTCTTCCAGAAAGTTCGCTCCGG 1200
QY 1200 CAGTGGGAGTCCCGTCAATAGTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTC 1259
DB 1201 CAGTGGGAGTCCCGTCAATAGTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTC 1260
QY 1260 GTCAAGAGATGAGCAGCAGCAGAGAAACCTGCTGTGACATTTGAAGATAAGAGCG 1319
DB 1261 GTCAAGAGATGAGCAGCAGCAGAGAAACCTGCTGTGACATTTGAAGATAAGAGCG 1320
QY 1320 GGAGAACTTTCAGAGGAGGAGTGTGAGCTGAGAGCGCGCCAGAGCTCTTGGAGCA 1379
DB 1321 GGAGAACTTTCAGAGGAGGAGTGTGAGCTGAGAGCGCGCCAGAGCTCTTGGAGCA 1380
QY 1380 GCAAGCGAAAGAGCAGAGCGGTTGGCTCAGCTGAGAGCGCGCGCGAGAGAGAAAGA 1439
DB 1381 GCAAGCGAAAGAGCAGAGCGGTTGGCTCAGCTGAGAGCGCGCGCGAGAGAGAAAGA 1440
QY 1440 GCGGAGGCGCAGAGAGAGGCGCAACGCGCAGCTGAGCTGAGAGAGCAGCTGAGAA 1499
DB 1441 GCGGAGGCGCAGAGAGAGGCGCAACGCGCAGCTGAGCTGAGAGAGCAGCTGAGAA 1500
QY 1500 GCAAGCGAGAGCTGAGAGCGGAGCGAGAGAGAGAGAGAGATGAGAGGCGCGCA 1559
DB 1501 GCAAGCGAGAGCTGAGAGCGGAGCGGAGAGAGAGAGAGAGATGAGAGGCGCGCA 1560
QY 1560 GGGCGGCAAAAGCGGAACTGAGAAAGCGCAACATTTGAATGAGAACGGAACCGGAGCA 1619
DB 1561 GGGCGGCAAAAGCGGAACTGAGAAAGCGCAACATTTGAATGAGAACGGAACCGGAGCA 1620
QY 1620 GGAATCTCTGAATCAGAGGAAACAAGAGAGAGGAGCGCGTGGCTGAGAGGCAAGAG 1679
DB 1621 GGAATCTCTGAATCAGAGGAAACAAGAGAGAGGAGCGCGTGGCTGAGAGGCAAGAG 1680
QY 1680 GAAAGCTTGTGAGTTGAGTTAGAACTCTGATGACAAAAAGCATCAGCTGAGAGAAA 1739
DB 1681 GAAAGCTTGTGAGTTGAGTTAGAACTCTGATGACAAAAAGCATCAGCTGAGAGAAA 1740
QY 1740 ACTTGAAGATATCAGGTGTGCACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAACA 1799
DB 1741 ACTTGAAGATATCAGGTGTGCACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAACA 1800
QY 1800 GTCTAGAGAGCTTAAGAAATTTGCTGAATACACCACCTTACAGCAGAGCTTGAGAAATCTCA 1859
DB 1801 GTCTAGAGAGCTTAAGAAATTTGCTGAATACACCACCTTACAGCAGAGCTTGAGAAATCTCA 1860
QY 1860 GCAAAATGCTTGAAGACTTATTCAGAGAAACAGATCTCAGTGAACAGTTAAAACAAGT 1919
DB 1861 GCAAAATGCTTGAAGACTTATTCAGAGAAACAGATCTCAGTGAACAGTTAAAACAAGT 1920
QY 1920 CCAAGCAAAACAGTTTGCATAGAGATCGCTTTTACCTCAAAAGAGCTTTGAAAGCAAA 1979
DB 1921 CCAAGCAAAACAGTTTGCATAGAGATCGCTTTTACCTCAAAAGAGCTTTGAAAGCAAA 1980
QY 1980 GGAAGCTGGCCGCGCAGAGCTCCGGGAGCAGCTGGAAGAGGTGAGAGAGAGCAAGGTC 2039
DB 1981 GGAAGCTGGCCGCGCAGAGCTCCGGGAGCAGCTGGAAGAGGTGAGAGAGAGCAAGGTC 2040
QY 2040 AAAAGCTGAGAGATGATGTTTTCACAAACAGCTGAAGAACTGAGAGAGATACATAG 2099
DB 2041 AAAAGCTGAGAGATGATGTTTTCACAAACAGCTGAAGAACTGAGAGAGATACATAG 2100
```

QY 2100 CAACAGCACTCCGAGAGAGAGTCCCTGAGGACGCGACTGAAGCAGAAAAGACA 2159
| | | | |
Db 2101 CAACAGCACTCCGAGAGAGAGTCCCTGAGGACGCGACTGAAGCAGAAAAGACA 2160
| | | | |
QY 2160 GGAAGAGAGAGCTTGAGATTAGAGAACAAAAGAGAGACGCTCAGAGACGATTCAAGA 2219
| | | | |
Db 2161 GGAAGAGAGAGCTTGAGATTAGAGAACAAAAGAGAGACGCTCAGAGACGATTCAAGA 2220
| | | | |
QY 2220 AAGGAGCAAGCAATGCTGAGACATGTCAGACAGAGAGACGACGCCCCCGGAAACC 2279
| | | | |
Db 2221 AAGGAGCAAGCAATGCTGAGACATGTCAGACAGAGAGACGACGCCCCCGGAAACC 2280
| | | | |
QY 2280 CCACGAGAGAGACAGACTGAAGAGGAGAGACAGTGTCAAGAAAGAGAGCGAGAGAG 2339
| | | | |
Db 2281 CCACGAGAGAGACAGACTGAAGAGGAGAGACAGTGTCAAGAAAGAGAGCGAGAGAG 2340
| | | | |
QY 2340 AGCCAAAGCCGGAATGCAAGACAAGCAGAGTGGCTTTTCATCCGATCAGAGCCAGC 2399
| | | | |
Db 2341 AGCCAAAGCCGGAATGCAAGACAAGCAGAGTGGCTTTTCATCCGATCAGAGCCAGC 2400
| | | | |
QY 2400 TAAAGTGGCCCAACCCGAGGACCCCTGGTCTTACACAGAGAAAGCCCGCTTACATTTCTGC 2459
| | | | |
Db 2401 TAAAGTGGCCCAACCCGAGGACCCCTGGTCTTACACAGAGAAAGCCCGCTTACATTTCTGC 2460
| | | | |
QY 2460 ACAGAGAGTGTAAAAGTGTATATTATTAACGAGCGCTGTACCCCTTTGAATCAGAAATCA 2519
| | | | |
Db 2461 ACAGAGAGTGTAAAAGTGTATATTATTAACGAGCGCTGTGTACCCCTTTGAATCAGAAATCA 2520
| | | | |
QY 2520 CGATGAGATCAACCATCCAGCCAGGAGATATAGTCATGTTGATGAAAGCAGACTGAGA 2579
| | | | |
Db 2521 CGATGAGATCAACCATCCAGCCAGGAGATATAGTCATGTTGATGAAAGCAGACTGAGA 2580
| | | | |
QY 2580 GCCAGAGTGGCTTGGAGAGAGCTGAAAAGGAGACGCGAGTGGTTCCCTGCAAACTATAGC 2639
| | | | |
Db 2581 GCCAGAGTGGCTTGGAGAGAGCTGAAAAGGAGACGCGAGTGGTTCCCTGCAAACTATAGC 2640
| | | | |
QY 2640 AGAAAAGATTCAGAAAATGAGAGTCCCACTCCAGCCAAACAGAGTCCGATCTGATC 2699
| | | | |
Db 2641 AGAAAAGATTCAGAAAATGAGAGTCCCACTCCAGCCAAACAGAGTCCGATCTGATC 2700
| | | | |
QY 2700 TGGCCCTGCCCCCAAACTGGCTCTGCGTAGAGACCCCTGCTCTTTGCCAGTGAACCTCTTC 2759
| | | | |
Db 2701 TGGCCCTGCCCCCAAACTGGCTCTGCGTAGAGACCCCTGCTCTTTGCCAGTGAACCTCTTC 2760
| | | | |
QY 2760 TGAAGCTTCACAAACCCCAACAACTGGGAGACTTCAAGTTCAGAGTGGCCCAAGAGTTC 2819
| | | | |
Db 2761 TGAAGCTTCACAAACCCCAACAACTGGGAGACTTCAAGTTCAGAGTGGCCCAAGAGTTC 2820
| | | | |
QY 2820 AAACGAGAGCCAGAAAAGGACAACTGGGATACGTGGGCGGCTCAGCCTTCTGAGCCGT 2879
| | | | |
Db 2821 AAACGAGAGCCAGAAAAGGACAACTGGGATACGTGGGCGGCTCAGCCTTCTGAGCCGT 2880
| | | | |
QY 2880 ACCTAGTCTGCGCCAGTTACGGCAGAGATCAGCCTTTTACCCGAGCAGACGCACTGGCTC 2939
| | | | |
Db 2881 ACCTAGTCTGCGCCAGTTACGGCAGAGATCAGCCTTTTACCCGAGCAGACGCACTGGCTC 2940
| | | | |
QY 2940 CTCCCACTCTCCCGTCTGCGGCGCAGGCTGAAAAGTGGAGAGGCTACAAAGCCCAAGCCCT 2999
| | | | |
Db 2941 CTCCCACTCTCCCGTCTGCGGCGCAGGCTGAAAAGTGGAGAGGCTACAAAGCCCAAGCCCT 3000
| | | | |
QY 3000 GTATCCCTGAGAGCCAAAAGAACAAACCACTTAATTTTAAACAAAAGTGAAGTATCAGC 3059
| | | | |
Db 3001 GTATCCCTGAGAGCCAAAAGAACAAACCACTTAATTTTAAACAAAAGTGAAGTATCAGC 3060
| | | | |
QY 3060 CGTTCCTGAGACAGCAAGACATGTGTGTTTGGAGAGATTCAAGGTCAAGAGGTTGGTT 3119
| | | | |
Db 3061 CGTTCCTGAGACAGCAAGACATGTGTGTTTGGAGAGATTCAAGGTCAAGAGGTTGGTT 3120
| | | | |
QY 3120 CCCCAGTCTTACGTGAATCTATTTCAAGGCGCTGAAGAAATCAACAGATCTGATAC 3179
| | | | |
Db 3121 CCCCAGTCTTACGTGAATCTATTTCAAGGCGCTGAAGAAATCAACAGATCTGATAC 3180
| | | | |
QY 3180 TGGGCTCTAGAAAGTCTGTGATGTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGC 3239
| | | | |

Db 3181 TGGGCTCTAGAAAGTCTGTGATGTAAAGAGAGTGGCTTCCCGGCGGCAAGCCAGC 3240
| | | | |
QY 3240 CATTCCCGGAGAGAGATTATTTGCGATGTACATACAGAGATTCTGACAGAGATTT 3299
| | | | |
Db 3241 CATTCCCGGAGAGAGATTATTTGCGATGTACATACAGAGATTCTGACAGAGATTT 3300
| | | | |
QY 3300 AACCTTCAAGCAAGGAGATGTATGTGTATCCAAAGAAAGATGATGATCTGTGAGAGCGG 3359
| | | | |
Db 3301 AACCTTCAAGCAAGGAGATGTATGTGTATCCAAAGAAAGATGATGATCTGTGAGAGCGG 3360
| | | | |
QY 3360 AACGTTGGGCGACAAAGTCCGAGTCTTCCCTTCACTATGTGAGCTTAAAGATTGAGA 3419
| | | | |
Db 3361 AACGTTGGGCGACAAAGTCCGAGTCTTCCCTTCACTATGTGAGCTTAAAGATTGAGA 3420
| | | | |
QY 3420 GGGCTCTGGAATCTCTGGAAAACAGGAGATTATGAAAAAAACCTGAATTTGCCAGGT 3479
| | | | |
Db 3421 GGGCTCTGGAATCTCTGGAAAACAGGAGATTATGAAAAAAACCTGAATTTGCCAGGT 3480
| | | | |
QY 3480 TATTGCTTCCAGCTGCTAGTCTGTCCGAAACAACCTGAGCTCTGGGCGAGTAT 3539
| | | | |
Db 3481 TATTGCTTCCAGCTGCTAGTCTGTCCGAAACAACCTGAGCTCTGGGCGAGTAT 3540
| | | | |
QY 3540 TCTGATCCGAAAAAGAACCCAGGTGATGTTGGGAAAGAGAACTGCAAGCTCGAGGAA 3599
| | | | |
Db 3541 TCTGATCCGAAAAAGAACCCAGGTGATGTTGGGAAAGAGAACTGCAAGCTCGAGGAA 3600
| | | | |
QY 3600 AAAGCCGCAATAGGAGTGTTCAGCAATATATGTCAAACTTTAAGCCCCGAGCAAG 3659
| | | | |
Db 3601 AAAGCCGCAATAGGAGTGTTCAGCAATATATGTCAAACTTTAAGCCCCGAGCAAG 3660
| | | | |
QY 3660 CAATAATCACCCCACTAGCTACCCAGACCCGAGTCCAGTCCAGTCCAGTAT 3719
| | | | |
Db 3661 CAATAATCACCCCACTAGCTACCCAGACCCGAGTCCAGTCCAGTCCAGTAT 3720
| | | | |
QY 3720 CGGATGTAGATTAACACCGCCAGAACGATGACGAATGACCTTTCAGCAAGCCAGAT 3779
| | | | |
Db 3721 CGGATGTAGATTAACACCGCCAGAACGATGACGAATGACCTTTCAGCAAGCCAGAT 3780
| | | | |
QY 3780 CATCAAGCTCTCAACAAAGAGACCCGAGTGTGAAAAGAGAAATCACTGTGGCAAGT 3839
| | | | |
Db 3781 CATCAAGCTCTCAACAAAGAGACCCGAGTGTGAAAAGAGAAATCACTGTGGCAAGT 3840
| | | | |
QY 3840 TGGGCTCTTCCCACTCAATTAATGTAAGTGAACCAAGACATGAGCCCAAGCCAGCAATG 3899
| | | | |
Db 3841 TGGGCTCTTCCCACTCAATTAATGTAAGTGAACCAAGACATGAGCCCAAGCCAGCAATG 3900
| | | | |

RESULT 3
AAZ39009
ID AAZ39009 standard; cDNA; 3723 BP.
XX
AC AAZ39009;
XX
AC
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse BseI coding sequence.
XX
KW Mouse; murine; BseI; Bse2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
OS Mus sp.
OS
PN W09955728-A2.
XX
PD 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
PF
XX 27-APR-1999; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI: 2000-052802/04.

XX P-PSDB; AAY57444.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection

XX Claim 6: Page 40-42; 99pp; English.

XX The present sequence encodes mouse Ese1. The present invention
CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
CC variants (Ese = Esi-domain and Sh3-domain regulator of endocytosis). (I)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Ese1 is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor;
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Eps15 complex, then binding dynamitin to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
XX
XX Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

Query Match 73.2%; Score 3723; DB 21; Length 3723;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 ATGGCTCAAGTTCACACACCTTCGGTGTAGCTGTGAGCCATACTGTGAG 318
DB 1 ATGGCTCAAGTTCACACACCTTCGGTGTAGCTGTGAGCCATACTGTGAG 60
QY 319 GAAAGGGCCAGCAGTACGAGCAGTTCCTTAGCTTGAAGCCGATGCGGATTTTACT 378
DB 61 GAAAGGGCCAGCAGTACGAGCAGTTCCTTAGCTTGAAGCCGATGCGGATTTTACT 120
QY 379 GGTGATCAAGCAGGAACTTTTTCCTCAATCTGGGTTCCTCAGCCTTCTTAGACAA 438
DB 121 GGTGATCAAGCAGGAACTTTTTCCTCAATCTGGGTTCCTCAGCCTTCTTAGACAA 180
QY 439 ATATGAGGCGCTAGCGGACATGATTAAGATGAGATGATCAAGTGAATTTTCCATA 498
DB 181 ATATGAGGCGCTAGCGGACATGATTAAGATGAGATGATCAAGTGAATTTTCCATA 240
QY 499 GCCATGAAGCTTATCAAACTGAAGCTACAAGATATCAGCTCCCTCCACACTTCCCTC 558
DB 241 GCCATGAAGCTTATCAAACTGAAGCTACAAGATATCAGCTCCCTCCACACTTCCCTC 300
QY 559 GTCAAGAAACAGCAACCAATGCTATTTCCAGTGCACCAAGCTTTGGTATAGAGGAGAT 618
DB 301 GTCAAGAAACAGCAACCAATGCTATTTCCAGTGCACCAAGCTTTGGTATAGAGGAGAT 360
QY 619 GGTAGATGCCACACTCAAGCTGTGCTCCGTGSCCAATGGGCTCCATTCCAGTTGT 678
DB 361 GGTAGATGCCACACTCAAGCTGTGCTCCGTGSCCAATGGGCTCCATTCCAGTTGT 420
QY 679 GGAATGTCTCAACCTTAGTATCTTGTCTCCTCCAGAGCAGTGTCTCCCTGTGCTAAC 738
DB 421 GGAATGTCTCAACCTTAGTATCTTGTCTCCTCCTCCAGAGCAGTGTCTCCCTGTGCTAAC 480

QY 739 GGGGCTCCTCCGTCATACAGCTGCTGAGTTCGGCATCTGACAGCCACATGGCCA 798
DB 481 GGGGCTCCTCCGTCATACAGCTGCTGAGTTCGGCATCTGACAGCCACATGGCCA 540
QY 799 AAGAGTTCCTCCTGACAGATCTGATCCAGGGTCACATTAAGAATACTAAGTTACAGAG 858
DB 541 AAGAGTTCCTCCTGACAGATCTGATCCAGGGTCACATTAAGAATACTAAGTTACAGAG 600
QY 859 GCAATATCTTGATGTGCCAGCGCCCTCCAGAGCAGAAATGGGCTGTGCTCAGTCA 918
DB 601 GCAATATCTTGATGTGCCAGCGCCCTCCAGAGCAGAAATGGGCTGTGCTCAGTCA 660
QY 919 TCAAGGCTGAATATAGGAGATTAATCAACACCCACGCAAAATATAGTGTGACACTTA 978
DB 661 TCAAGGCTGAATATAGGAGATTAATCAACACCCACGCAAAATATAGTGTGACACTTA 720
QY 979 ACAGGTCCCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 1038
DB 721 ACAGGTCCCCAGGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780
QY 1039 TCAATATGGAATCTTTTGCACATTGATCAAGTGAATACTCAGTCGAGAGAAATTTATC 1098
DB 781 TCAATATGGAATCTTTTGCACATTGATCAAGTGAATACTCAGTCGAGAGAAATTTATC 840
QY 1099 CTAGCTATGACCTAATTAATGATGTCATGCTGTCAGCCACCTGCGCCCTGCTCCT 1158
DB 841 CTAGCTATGACCTAATTAATGATGTCATGCTGTCAGCCACCTGCGCCCTGCTCCT 900
QY 1159 CCAGAAATCATCT 1218
DB 901 CCAGAAATCATCT 960
QY 1219 AGCTCTTCT 1278
DB 961 AGCTCTTCT 1020
QY 1279 CCAGAGAAAGAACTGCTGTGATTTGAAGTAAAGACCGGAGAACTTCCAGAGAGCC 1338
DB 1021 CCAGAGAAAGAACTGCTGTGATTTGAAGTAAAGACCGGAGAACTTCCAGAGAGCC 1080
QY 1339 AGTGTGAGCTGTGAAGAGCGCCGCAAGCGCTCTTGAAGCAGCAGCGCAAGAGAGAG 1398
DB 1081 AGTGTGAGCTGTGAAGAGCGCCGCAAGCGCTCTTGAAGCAGCAGCGCAAGAGAGAG 1140
QY 1399 CGGTGGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1458
DB 1141 CGGTGGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1459 GAGGCTCAAGCGGACACTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
DB 1201 GAGGCTCAAGCGGACACTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1519 CAGCAG 1578
DB 1261 CAGCAG 1320
QY 1579 GAAAGGACGCAAACTTGAATGGAAACCGAAACCGAGACAGAACTCCTGAATCAGAGG 1638
DB 1321 GAAAGGACGCAAACTTGAATGGAAACCGAAACCGAGACAGAACTCCTGAATCAGAGG 1380
QY 1639 AACAG 1698
DB 1381 AACAG 1440
QY 1699 TTAGAAGCTCTGAATGAACAAAAGATCAGTAAAGAGAAAACCTCAGAGATACAGGTGT 1758
DB 1441 TTAGAAGCTCTGAATGAACAAAAGATCAGTAAAGAGAAAACCTCAGAGATACAGGTGT 1500
QY 1759 CGACTGGCAACCCAGAGGCAAGAAATTAAGAGCAGAAACAAGCTTAAGAGCTAAGAAAT 1818
DB 1501 CGACTGGCAACCCAGAGGCAAGAAATTAAGAGCAGAAACAAGCTTAAGAGCTAAGAAAT 1560

QY 1819 GCTGAATACCCCACTTACAGACAGAGTTCAGAAATCTCAGCAAAATGCTTGAAGACTT 1878
| | | | |
Db 1861 GCTGAATACCCCACTTACAGACAGAGTTCAGAAATCTCAGCAAAATGCTTGAAGACTT 1820
| | | | |
QY 1879 ATTCAGAGAAACAGATATCTAGTGCAGATTAAAAAAGTTCAGCAGAAAGATTGGAT 1938
| | | | |
Db 1621 ATTCAGAGAAACAGATATCTAGTGCAGATTAAAAAAGTTCAGCAGAAAGATTGGAT 1680
| | | | |
QY 1939 AGAGATGCTTCTTAAACCTCAAAAGACCTTGAAGCAAGAGAGTGGCCCGCAGAG 1998
| | | | |
Db 1681 AGAGATGCTTCTTAAACCTCAAAAGACCTTGAAGCAAGAGAGTGGCCCGCAGAG 1740
| | | | |
QY 1999 CTCGCGAGCAGCTTCAGAGGTGAGAGAGACCAAGTCAAAAGCTCAGAGATTTGAT 2058
| | | | |
Db 1741 CTCGCGAGCAGCTTCAGAGGTGAGAGAGACCAAGTCAAAAGCTCAGAGATTTGAT 1800
| | | | |
QY 2059 GTTTTCAACAAACAGCTGAAGAACTGAGAGATATACATGCAAAACAGCAATCCAGAG 2118
| | | | |
Db 1801 GTTTTCAACAAACAGCTGAAGAACTGAGAGATATACATGCAAAACAGCAATCCAGAG 1860
| | | | |
QY 2119 CAGAGTCCCTGAGAGGAGCCGCACTGAAGCAAGAAAGCAGAGAGAGAGAGCTTGAG 2178
| | | | |
Db 1861 CAGAGTCCCTGAGAGGAGCCGCACTGAAGCAAGAAAGCAGAGAGAGAGAGCTTGAG 1920
| | | | |
QY 2179 TTAGAAGCAAAAGAGAGAGAGCTCAGAGAGAGTTCAGAAAGGGAACAAGATGGCTG 2238
| | | | |
Db 1921 TTAGAAGCAAAAGAGAGAGAGCTCAGAGAGAGTTCAGAAAGGGAACAAGATGGCTG 1980
| | | | |
QY 2239 GAGCATGTGACAGAGAGAGAGAGCAGCCGCCGGAACCCCAAGAGAGAGAGAGCTG 2298
| | | | |
Db 1981 GAGCATGTGACAGAGAGAGAGAGCAGCCGCCGGAACCCCAAGAGAGAGAGAGCTG 2040
| | | | |
QY 2299 AAGAGGGAAGACAGTGTAGAGAAAGAGAGGCGGAAGAGAGAGAGAGAGAGAGAGAG 2358
| | | | |
Db 2041 AAGAGGGAAGACAGTGTAGAGAAAGAGAGGCGGAAGAGAGAGAGAGAGAGAGAGAG 2100
| | | | |
QY 2359 GACAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2418
| | | | |
Db 2101 GACAAGCAGAGTGGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGGCCACCCAGGCA 2160
| | | | |
QY 2419 CCTGTGTACCAAGAGAAAGCCCGCTTACATTTCTGCAAGAGAGAGTGTAAAGTG 2478
| | | | |
Db 2161 CCTGTGTACCAAGAGAAAGCCCGCTTACATTTCTGCAAGAGAGAGTGTAAAGTG 2220
| | | | |
QY 2479 GTATATTTACGAGCGCTGTACCCCTTTGAATCCAGAAATTCAGATGAGATACCATCCAG 2538
| | | | |
Db 2221 GTATATTTACGAGCGCTGTACCCCTTTGAATCCAGAAATTCAGATGAGATACCATCCAG 2280
| | | | |
QY 2539 CCAGAGATATATGTCATGCTGATGAAGCAAGCTGAGAGCCAGATGGCTTGAGAGA 2598
| | | | |
Db 2281 CCAGAGATATATGTCATGCTGATGAAGCAAGCTGAGAGCCAGATGGCTTGAGAGA 2340
| | | | |
QY 2599 GAGCTGAAGGGAAGACGGGATGGTTCCTGCAAACTATGCAAGAAAGATTCAGAAAT 2658
| | | | |
Db 2341 GAGCTGAAGGGAAGACGGGATGGTTCCTGCAAACTATGCAAGAAAGATTCAGAAAT 2400
| | | | |
QY 2659 GAGGTTCCTCACTCCAGCCAAACCAAGTGAACGATGCAATCTGCCCCCTCCCAATG 2718
| | | | |
Db 2401 GAGGTTCCTCACTCCAGCCAAACCAAGTGAACGATGCAATCTGCCCCCTCCCAATG 2460
| | | | |
QY 2719 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTTCAACAACCCC 2778
| | | | |
Db 2461 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTTCAACAACCCC 2520
| | | | |
QY 2779 AACAACTGGGAGACTTCAGTTCCAGTGGCCAGCAGCTCAAAAGAGAGCCAGAAACG 2838
| | | | |
Db 2521 AACAACTGGGAGACTTCAGTTCCAGTGGCCAGCAGCTCAAAAGAGAGCCAGAAACG 2580
| | | | |
QY 2839 GACAACATGGGATATGCTGGGCGGCTCAGCTTCTGAGACCTTACCTAAGTGGCCAGTTA 2898
| | | | |
Db 2581 GACAACATGGGATATGCTGGGCGGCTCAGCTTCTGAGACCTTACCTAAGTGGCCAGTTA 2640
| | | | |
QY 2899 CGGAGAGATAGGCTTTACCCAGCAGCAGCAGCTGCTCCTCCATCTCCGCTCTG 2958
| | | | |

Db 2641 CGGAGAGATAGGCTTTACCCAGCCAGCAGCTGCTCTCCCATCTCCGCTCTG 2700
| | | | |
QY 2959 GGCCAGGGTGAAGAGTGGAAAGGGCTTCAAGCCGAAACCTCTGTATCTCTGAGAGCCAAA 3018
| | | | |
Db 2701 GGCCAGGGTGAAGAGTGGAAAGGGCTTCAAGCCGAAACCTCTGTATCTCTGAGAGCCAAA 2760
| | | | |
QY 3019 AAAGACAACCACTTAATTTTAAACAAAGTGAAGCTCATCACCGTTCTGGAACAGCAAGAC 3078
| | | | |
Db 2761 AAAGACAACCACTTAATTTTAAACAAAGTGAAGCTCATCACCGTTCTGGAACAGCAAGAC 2820
| | | | |
QY 3079 ATGTGTGTGTTTGGAGAGTTCAAGTCAAGAGGGTGTGTTCCCAAGTCTTACGTGAAA 3138
| | | | |
Db 2821 ATGTGTGTGTTTGGAGAGTTCAAGTCAAGAGGGTGTGTTCCCAAGTCTTACGTGAAA 2880
| | | | |
QY 3139 CTCATTTTCAAGGCCCCGTAAAGAAATTCACACAGCATTCATCTGCCCCCTTACGAAAGTCT 3198
| | | | |
Db 2881 CTCATTTTCAAGGCCCCGTAAAGAAATTCACACAGCATTCATCTGCCCCCTTACGAAAGTCT 2940
| | | | |
QY 3199 GCTAGTCTAAAGAGAGTGGCTTCCCGCGCCCAAGCAGCAGTCTCCCGAGAAAGTTT 3258
| | | | |
Db 2941 GCTAGTCTAAAGAGAGTGGCTTCCCGCGCCCAAGCAGCAGTCTCCCGAGAAAGTTT 3000
| | | | |
QY 3259 ATTCGATGTACACATACGAGAGTTCTGAGCAAGAGATTTAACCTTTCAAGCAAGGAGAT 3318
| | | | |
Db 3001 ATTCGATGTACACATACGAGAGTTCTGAGCAAGAGATTTAACCTTTCAAGCAAGGAGAT 3060
| | | | |
QY 3319 GTGATTTGTGTTTACCAAGAAAGATGTGATCTGTGACCGGAAACGTGTGGGACAAAGTCC 3378
| | | | |
Db 3061 GTGATTTGTGTTTACCAAGAAAGATGTGATCTGTGACCGGAAACGTGTGGGACAAAGTCC 3120
| | | | |
QY 3379 GGAATCTTCCCTTTAACTTATGTAGAGCTTAAAGATTCAAGAGGCTCTGAAATGCTGGG 3438
| | | | |
Db 3121 GGAATCTTCCCTTTAACTTATGTAGAGCTTAAAGATTCAAGAGGCTCTGAAATGCTGGG 3180
| | | | |
QY 3439 AAAACAGGAGTTTGAAGAAAAAACTGAATTGCCCCAGGTTATGCTTCTTACGCTGCT 3498
| | | | |
Db 3181 AAAACAGGAGTTTGAAGAAAAAACTGAATTGCCCCAGGTTATGCTTCTTACGCTGCT 3240
| | | | |
QY 3499 ACTGCTCCCGAACAATCAATCCCTGCTCTGGGACGCTGATTCGTATCCGAAAAAAGAAC 3558
| | | | |
Db 3241 ACTGCTCCCGAACAATCAATCCCTGCTCTGGGACGCTGATTCGTATCCGAAAAAAGAAC 3300
| | | | |
QY 3559 CCAGGTGAGTGTGGGAAGGAACTGCAAGCTGAGAGGAAAAAGCCAGATAGGGTGG 3618
| | | | |
Db 3301 CCAGGTGAGTGTGGGAAGGAACTGCAAGCTGAGAGGAAAAAGCCAGATAGGGTGG 3360
| | | | |
QY 3619 TTTTCAGCAAAATTAATGTCAAACTTCTAAGCCCGGAAACAAGCAAAATCACCCCAATGAG 3678
| | | | |
Db 3361 TTTTCAGCAAAATTAATGTCAAACTTCTAAGCCCGGAAACAAGCAAAATCACCCCAATGAG 3420
| | | | |
QY 3679 CTACCCAAAGCCGAGTGCAGCCAGCAGTGTGCCAGGTGATCGGAGTGTACATTAACC 3738
| | | | |
Db 3421 CTACCCAAAGCCGAGTGCAGCCAGCAGTGTGCCAGGTGATCGGAGTGTACATTAACC 3480
| | | | |
QY 3739 GCCCAGAACGATGAGAACTAGCTTACAGCAAGGCGAGATCATCAAGTCTTCAACAG 3798
| | | | |
Db 3481 GCCCAGAACGATGAGAACTAGCTTACAGCAAGGCGAGATCATCAAGTCTTCAACAG 3540
| | | | |
QY 3799 GAGGACCCGAGCTGTGGAAGAGAGAGTCAAGTGGGCAAGTGGGCTTTCCCATTCAGAT 3858
| | | | |
Db 3541 GAGGACCCGAGCTGTGGAAGAGAGAGTCAAGTGGGCAAGTGGGCTTTCCCATTCAGAT 3600
| | | | |
QY 3859 TATGTAAGCTGACACAGACATGAGACCCAGCCAGCAAGATGATATGTTGTTCATCC 3918
| | | | |
Db 3601 TATGTAAGCTGACACAGACATGAGACCCAGCCAGCAAGATGATATGTTGTTCATCC 3660
| | | | |
QY 3919 CCCCCTCAGGCTTGAAGTCCCTCAAGAGCCCATATCCATATCACTGCCAGAGGGA 3978
| | | | |
Db 3661 CCCCCTCAGGCTTGAAGTCCCTCAAGAGCCCATATCCATATCACTGCCAGAGGGA 3720
| | | | |
QY 3979 TGA 3981
| | | | |

Db 3721 TGA 3723

RESULT 4
AAZ39025
ID AAZ39025 standard; cDNA; 5144 BP.

XX AAZ39025;
XX
XX
XX 28-FEB-2000 (first entry)
XX
XX Mouse Esell coding sequence.
XX
XX
XX Mouse; murine; Esel1; Esel2; endocytosis; vesicular trafficking;
XX regulation; actin cytoskeleton; detection; cancer; infection;
XX EH-domain and SH3-domain regulator of endocytosis; anticancer;
XX antiproliferative; antiviral; ss.
XX
XX Mus sp.
XX
XX WO955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
XX
XX MPI: 2000-052802/04.
XX P-PSDB; AAY57449.
XX
XX New nucleic acid encoding Esel and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection -
XX
XX Claim 6; Page 59-62; 99pp; English.

The present invention specifically describes mammalian Esel and 2 proteins (I) and their splice variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I) are involved in regulation of clathrin-mediated endocytosis (as a complex with Eps15 protein), vesicular trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists, mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab); sequences antisense to the (I) polynucleotide; agents that downregulate expression of Ese genes or antagonists of an Ese binding partner are used to treat diseases associated with undesirable endocytosis and resulting changes in cellular function. Particularly overexpression of Esel is used to block clathrin-mediated endocytosis in vivo or in cell cultures, while administration of (I) is used to promote endocytosis of selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of cells that can be stimulated to proliferate by a growth factor receptor; and similar compounds (also inactive Ese mutants) can be used to prevent viral infection. Endocytosis may also be regulated, in vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding dynamin to the complex. Generally conditions that can be treated include cancer; abnormal cell division or migration; viral infection; or abnormal receptor signaling, tissue development or synaptic transmission. The present sequence represents mouse Esell coding sequence.

Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;

Query Match 71.6%; Score 3639.4; DB 21; Length 5144;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTCAAGTTTCCACACCTTTGGGTGAGCTGATGTCGGGCCATTAAGTGTGAG 318
DB 1 ATGGCTCAAGTTTCCACACCTTTGGGTGAGCTGATGTCGGGCCATTAAGTGTGAG 60

QY 319 GAAAGGGCCAAAGCATGACACAGAGTTCCCTTAGCCCTGAAGCCGATAGCGGGATTTTATTACT 378
DB 61 GAAAGGGCCAAAGCATGACACAGAGTTCCCTTAGCCCTGAAGCCGATAGCGGGATTTTATTACT 120

QY 379 GGTGATCAAGCAGGAACTTTTTCATCATCTGGGTTACTCAGCTGTCTTACACAA 438
DB 121 GGTGATCAAGCAGGAACTTTTTCATCATCTGGGTTACTCAGCTGTCTTACACAA 180

QY 439 ATATGGGGCTTAGCGGACATGAATACATGAGTAAGATGATCAAGTGAATTTTCCATA 498
DB 181 ATATGGGGCTTAGCGGACATGAATACATGAGTAAGATGATCAAGTGAATTTTCCATA 240

QY 499 GCCATGAAGCTTATCAACTGAAGTACAGGATATCAGCTCCCTCCACACTTCCCTCCCT 558
DB 241 GCCATGAAGCTTATCAACTGAAGTACAGGATATCAGCTCCCTCCACACTTCCCTCCCT 300

QY 559 GTCATGAACACAGCAACAGTGGCTATTTCCAGTGACACAGCATTTGTTATAGAGAGAT 618
DB 301 GTCATGAACACAGCAACAGTGGCTATTTCCAGTGACACAGCATTTGTTATAGAGAGAT 360

QY 619 GCTAGCATGCCACCACTACAGCTGTGCTCTGTGCAATGGGCTCCATTCCAGTTGTT 678
DB 361 GCTAGCATGCCACCACTACAGCTGTGCTCTGTGCAATGGGCTCCATTCCAGTTGTT 420

QY 679 GGAATGTCACCCCTTGTATCTCTGTCCTCCAGAGAGAGAGGCTCCCTGGGCTAAC 738
DB 421 GGAATGTCACCCCTTGTATCTCTGTCCTCCAGAGAGAGAGGCTCCCTGGGCTAAC 480

QY 739 GGGGCTCCTCCCGTATCAGGCTGTGCTGGCTGGGATCCGTCAGCCACATAGGCCA 798
DB 481 GGGGCTCCTCCCGTATCAGGCTGTGCTGGCTGGGATCCGTCAGCCACATAGGCCA 540

QY 799 AAGGTTCTTCTTACAGATCTGCTCAGGCTCAATTAACACTTAAGTTACAGAAG 858
DB 541 AAGGTTCTTCTTACAGATCTGCTCAGGCTCAATTAACACTTAAGTTACAGAAG 600

QY 859 GCACATCATTCGATGTGGCGGCGCCCTCCAGAGAGAGATGGGCTGTGCTCAGTCA 918
DB 601 GCACATCATTCGATGTGGCGGCGCCCTCCAGAGAGAGATGGGCTGTGCTCAGTCA 660

QY 919 TCAAGGCTGAATATACAGGCAAGTTATTAACAGCCACGACAAACTATGATGAGACACTTA 978
DB 661 TCAAGGCTGAATATACAGGCAAGTTATTAACAGCCACGACAAACTATGATGAGACACTTA 720

QY 979 ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGCT 1038
DB 721 ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGCT 780

QY 1039 TCAATATGGAATCTTTGACATTCATGATCAAGTGAAGAACTCATGCGAGAAATTTATTC 1098
DB 781 TCAATATGGAATCTTTGACATTCATGATCAAGTGAAGAACTCATGCGAGAAATTTATTC 840

QY 1099 CTAGCTATGACCTTATGATGATGTTGCCATGTCTGTGTCAGCCACTGCGCGCTCTGCT 1158
DB 841 CTAGCTATGACCTTATGATGATGTTGCCATGTCTGTGTCAGCCACTGCGCGCTCTGCT 900

QY 1159 CCAGAAATCATCCCTCTCTTCCAGAAAGTTGCTCCGAGTGGAGATGTCGTCATA 1218
DB 901 CCAGAAATCATCCCTCTCTTCCAGAAAGTTGCTCCGAGTGGAGATGTCGTCATA 960

QY 1219 AGCTTTCTTCTGTGATCAAGAGCTGCTGAGAGCCGCTGCTCAGAGATGAGACAGAG 1278
DB 961 AGCTTTCTTCTGTGATCAAGAGCTGCTGAGAGCCGCTGCTCAGAGATGAGACAGAG 1020

QY 1279 CCAGAGAGAAACTCTGTGATTCATTTGAAGTAAGAGCGGAGAACTTCAGAGGAGGC 1338
DB 1021 CCAGAGAGAAACTCTGTGATTCATTTGAAGTAAGAGCGGAGAACTTCAGAGGAGGC 1080

QY 1339 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAAGACAGACGCAAGAGCAGAG 1398
DB 1081 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAAGACAGACGCAAGAGCAGAG 1140

1399 CGTTTGCTCAGCTGAGCGCCGAGCAGAGAGAGAAAGACGGGACCGCCAGAGAG 1458
1141 CGGTTGGCTCAGCTGAGCGCCGAGCAGAGAGAGAAAGACGGGACCGCCAGAGAG 1200
1459 GAGGCAAGCGGACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1518
1201 GAGGCAAGCGGACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1519 CAGGAG 1578
1261 CAGCAG 1320
1579 GAAAGGCAAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1638
1321 GAAAGGCAAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
1639 AACAG 1698
1381 AACAG 1440
1699 TTAAGAGCTTGAATGACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758
1441 TTAAGAGCTTGAATGACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1759 GCACTGGCAACCCAGAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAGAG 1818
1501 GCACTGGCAACCCAGAGGCAAGAAATTTAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
1819 GCTGAATCACCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
1561 GCTGAATCACCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
1879 ATTCAG 1938
1621 ATTCAG 1680
1939 AGAGAGCTGCTTCTTAACTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1998
1681 AGAGAGCTGCTTCTTAACTCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
1999 CTCGGGAG 2058
1741 CTCGGGAG 1800
2059 GTTTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2118
1801 GTTTTCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
2119 CAGAGTCCCTGAG 2178
1861 CAGAGTCCCTGAG 1920
2179 TTAAG 2238
1921 TTAAG 1980
2239 GAGCATGTGCAAG 2298
1981 GAGCATGTGCAAG 2040
2299 AAG 2358
2041 AAG 2100
2359 GACAG 2418
2101 GACAG 2160
2419 CCGTGTGCTACACAG 2478
2161 CCGTGTGCTACACAG 2220
2479 GTATATTACCGAG 2538

2221 GTATATTACCGAG 2280
2539 CCAGAGATATAGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2598
2281 CCAGAGATATAGTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
2599 GAGCTGAAG 2658
2341 GAGCTGAAG 2400
2659 GAGGTTCCCACTCCAGGCAAAACAGTGAACGATCTGATCTGACCTGCCCCAAACTG 2718
2401 GAGGTTCCCACTCCAGGCAAAACAGTGAACGATCTGATCTGACCTGCCCCAAACTG 2460
2719 GCTTGTGCTGAG 2778
2461 GCTTGTGCTGAG 2520
2779 AACAACTGGGAG 2838
2521 AACAACTGGGAG 2580
2839 GACAACTGGGAG 2898
2581 GACAACTGGGAG 2640
2899 CGGAGAGATATAGGCTTTTACCCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2958
2641 CGGAGAGATATAGGCTTTTACCCAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
2959 GGGCAG 3018
2701 GGGCAG 2760
3019 AAGAGCAACCACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAG 3078
2761 AAGAGCAACCACTTAAATTTTAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
3079 ATGTGTGTGTTTGAAG 3138
2821 ATGTGTGTGTTTGAAG 2880
3139 CTCATTTCAAGGAG 3198
2881 CTCATTTCAAGGAG 2940
3199 GCTAGCTTAAAG 3258
2941 GCTAGCTTAAAG 3000
3259 ATTGCAATGTACATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3318
3001 ATTGCAATGTACATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
3319 GTGATGTGTGTACCAAG 3378
3061 GTGATGTGTGTACCAAG 3120
3379 GGAATCTTCCCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3438
3121 GGAATCTTCCCTTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
3439 AAAACAGGAG 3498
3181 AAAACAGGAG 3240
3499 ACTGATCCCGAACAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3558
3241 ACTGATCCCGAACAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
3559 CCAAGTGTAGTGTGAG 3618


```

Db      3301 CCAAGTGTAGTGGGAGAGAGAACTCGAAGTGGAGGAAAAAGCCGACGATGAGGTGG 3360
Qy      3619 TTTCAGCAAAATATATCTTAAAGTCCCGGAAAGCAAGCAAAATCACCCACTGAG 3678
Db      3361 TTTCAGCAAAATATATCTTAAAGTCCCGGAAAGCAAGCAAAATCACCCACTGAG 3420
Qy      3679 CTAACCAAGACCGCACTGACGACGAGTGTGCGAGGTATCGGATGTACATTAACACC 3738
Db      3421 CTAACCAAGACCGCACTGACGACGAGTGTGCGAGGTATCGGATGTACATTAACACC 3480
Qy      3739 GCCCAAGACGATGACCACTAGCTTACAGCAAAAGCCCAATCATCACTCTCAACAG 3798
Db      3481 GCCCAAGACGATGACCACTAGCTTACAGCAAAAGCCCAATCATCACTCTCAACAG 3540
Qy      3799 GAGACCCGCACTGTGTGAAAGAGAGTCACTGAGGCAATGTTCCATCCAT 3858
Db      3541 GAGACCCGCACTGTGTGAAAGAGAGTCACTGAGGCAATGTTCCATCCAT 3600
Qy      3859 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCAATG 3899
Db      3601 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCAATG 3641

RESULT 5
AAZ34571
ID      AAZ34571 standard; cDNA; 5458 BP.
AC      AAZ34571;
DT      01-FEB-2000 (first entry)
DE      Human SH3D1A cDNA clone 21.
KW      SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
        megakaryocytic abnormality; myeloproliferative disorder;
        platelet disorder; neural disorder; thrombocytopenia;
        haematopoietic disorder; cognitive dysfunction; microcephaly;
        lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
        ss.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      CDS      267..3929
          FT      /tag= a
          FN      WO9953062-A2.
          PD      21-OCT-1999.
          PF      16-APR-1999; 99WO-US08371.
          PR      16-APR-1998; 98US-0082007.
          PA      (CEDA-) CEDARS SINAI HEALTH SYSTEM.
          PI      Korenberg JR, Chen X;
          DR      WPI; 1999-633829/54.
          DX      P-PSDB; AAY32155.
          PT      Nucleic acid from the human SH3D1A gene and its products, useful for
          PT      the diagnosis and treatment of myeloproliferative disorders and
          PT      leukaemia.
          PS      Claim 2; Fig 8; 99p; English.
          CC      This is the nucleotide sequence of full-length cDNA (clone 21)
          CC      corresponding to a novel human SH3 gene, termed the SH3D1A gene,
          CC      that contributes to the development of platelets and the
          CC      pathogenesis of leukaemias, both in general and in particular those
          CC      involving the megakaryocytic lineage. The SH3D1A gene maps to the
          CC      small candidate region for low platelets on chromosome 21.

```

```

CC      Sequencing of 5 different sizes of cDNA clone from foetal brain
CC      (see AAZ34570-74) suggests that at least 3 isoforms exist. The
CC      invention provides methods for the diagnosis and treatment of
CC      megakaryocytic abnormality, myeloproliferative disorder, platelet
CC      disorder, acute leukaemia, neural disorders, thrombocytopenia,
CC      platelet disorder on chromosome 21, low platelets in deletion for
CC      21, association of gains in chromosome 21 with leukaemias, neural
CC      abnormalities, dysfunctions and disorders including brain
CC      malformations and corresponding cognitive dysfunctions,
CC      microcephaly, lissencephaly, and colpocephaly. Methods are also
CC      provided for: suppressing cells unable to regulate themselves;
CC      screening for a somatic alteration in the SH3D1A gene; monitoring
CC      the progress and adequacy of a treatment; monitoring tumour risk
CC      progress or megakaryocytic abnormality, myeloproliferative disorder,
CC      haematopoietic disorder, platelet disorder or leukaemia; and
CC      treatment of a subject (including a prenatal subject) having
CC      megakaryocytic abnormality, myeloproliferative disorder,
CC      platelet disorder, leukaemia or neural disorder using a
CC      nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX      Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 other:

Query Match      69.4%; Score 3529.4; DB 20; Length 5458;
Best Local Similarity 83.7%; Pred. No. 0;
Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;

Qy      7 GAGGAGAGTGAAGCGCGCGGAGGCGCGGAGCTTGCTGCTGCTGCTGCTGCTGCTGCT 66
Db      25 GAGGAGAGTGAAGCGCGCGGAGGCGCGGAGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 84
Qy      67 CGGAGGAGGAGTCCCGAGCGGCGCTCCGGGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 126
Db      85 CGGAGGAGGAGTCCCGAGCGGCGCTCCGGGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 133
Qy      127 GGGATGTGTGTGCGCGCGCTGCGGACTCGGCGTTCCTGCG-GCGGCGGTGCGGCGTGACTGA 185
Db      134 GGGATGTGTGTGCGCGCGCTGCGGCTCTGCGTCTCTCCAGCGCGCGGTGAGGCGGACTGA 193
Qy      186 TTTGTGTGAGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 245
Db      194 TTTGTCTCCGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253
Qy      246 ACGTAATGAACATGAGGCTCACTTCCACACCTTCCGTTAGCTGAGTGTGCTGCGGCG 305
Db      254 AAGTAACAGAACATGAGGCTCACTTCCACACCTTTCGTCAGCTGAGTGTGCTGCGGCG 313
Qy      306 CATTAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 365
Db      314 CATTAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 373
Qy      366 GGGATTTTATCTGCTGATCAAGCGAGAACTTTTTCATCTGCGTTACTGAGCC 425
Db      374 TGGATTTTATCTGCTGATCAAGCTGAAGAACTTTTTCATCTGCGTTACTGAGCC 433
Qy      426 TGTCTTAGACAAATATGGGCGCTAGCGGAGCAATGAATGATGATGATGATGATGATGATGAT 485
Db      434 TGTCTTAGACAAATATGGGCGCTAGCGGAGCAATGAATGATGATGATGATGATGATGATGAT 493
Qy      486 GGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGTACAAAGATATCAGCTCCCTC 545
Db      494 GGAATTTTTCATAGCCATGAAAGCTTATCAAACTGAAGTACAAAGATATCAGCTCCCTC 553
Qy      546 CACACTTCCCTCTGTCATGAAAGAGCAACCAAGTGCATTTTCAGTGCACGCAATTTGG 605
Db      554 TGCACCTTCCCTCTGTCATGAAAGAGCAACCAAGTGCATTTTCAGTGCACGCAATTTGG 613
Qy      606 TATAGAGGAGTGTAGATGCGACCACTACAGCTGTGCTCTGTCGCAATGGGCTC 665
Db      614 TATAGAGGATATGCGACAGATGCGACCGCTTACAGCTGTGCTCTGTCGCAATGGGATC 673
Qy      666 CATTCAGTGTGTGAATGTCTCAACCTTAGATATCTGTGCTCTCAGCAGCAGAGTGC 725
Db      674 CATTCAGTGTGTGAATGTCTCAACCTTAGATATCTGTGCTCTCAGCAGCAGAGTGTGC 733

```


726 TCCCTGGCTAAGCGGGGCTCCCTCCCGTGAACAGCCTGCGCTTGGCATCCGCG 785
734 CCCCCGTGCTAAGCGGGGCTCCCGCTGTATACACCTGCGCTGATTTGCTCATCTGC 793
786 AGCCACATGCGCAAGAGTCTTCTTCAGAGATCTGGTCCAGGGTCACATTAACAC 845
794 AGCCACATGCGCAAGAGTCTTCTTCAGAGATCTGGTCCAGGGTCACATTAACAC 853
846 TAAATTACAAAGGACAGTCAATTTGATGTGGCCAGTGTCCACAGTGGGAGTGGGC 905
854 TAAATTACAAAGGACAGTCAATTTGATGTGGCCAGTGTCCACAGTGGGAGTGGGC 913
906 TGTGCTCAGTCAATCAAGGCTGAAATACAGGAGTATTCAACAGCCGCAAAACTAT 965
914 TGTTCCTCAGTCAATCAAGGCTGAAATACAGGAGTATTCAATGCTCAACAAACTAT 973
966 GAGTGAACCTTAAACAGGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCCA 1025
974 GAGTGAACCTTAAACAGGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCCA 1033
1026 GAGTGAACCTTAAACAGGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCCA 1085
1034 GAGTGAACCTTAAACAGGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCCA 1093
1086 AGAAGAAATTTATCCAGTATGACACCTAATGATGTTGGCATGCTGCTGACCACTGCC 1145
1094 AGAAGAAATTTATCCAGTATGACACCTAATGATGTTGGCATGCTGCTGACCACTGCC 1153
1146 GCGCGTCTGCTCCAGAAATACATCCCTCTTCTTCAGAAAGATTCGCTCCGAGTGG 1205
1154 ACCTGCTCTGCTCCAGAAATACATCCCTCTTCTTCAGAAAGATTCGCTCCGAGTGG 1213
1206 GATGTCCTCATTAAGCTCTTCTTCTGATGATCAGAGCTGCTGAGGAGCTGCTGCA 1265
1214 TATATCTGCTCATTAAGCTCTTCTTCTGATGATCAGAGCTGCTGAGGAGGAGGAGT 1273
1266 GAGTGAACCTTAAACAGGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCCA 1322
1274 AGAAGAAATTTATCCAGTATGACACCTAATGATGTTGGCATGCTGCTGACCACTGCC 1333
1323 GAACTTCAGAGGAGGACAGTGTGAGCTGAGAGAGGCGCGCAAGGCTCTTTCAGACACA 1382
1334 GAACTTCAGAGGAGGACAGTGTGAGCTGAGAGAGGCGCGCAAGGCTCTTTCAGACACA 1393
1383 GCGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1442
1394 GCGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1453
1443 GAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1502
1454 TGAAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1513
1503 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1562
1514 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1573
1563 GCGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1622
1574 TGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1633
1633 ACTGCTGAATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1682
1634 ACTGCTGAATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1693
1683 GACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1742
1694 GACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1753
1743 TCAAGATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1802
1754 TCAAGATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1813

1803 TAGAGAGTAAAGATGCTGAAATCAACCACTTACAGCAGAGGATGAGAGATCTCAGCA 1862
1814 TAGAGAGTAAAGATGCTGAAATCAACCACTTACAGCAGAGGATGAGAGATCTCAGCA 1873
1863 AATGCTTGAAGAGCTTATTTCCAGAGAAACAGATATCTAGTACCAAGTTTAAACAAAGTCCA 1922
1874 AATGCTTGAAGAGCTTATTTCCAGAGAAACAGATATCTAGTACCAAGTTTAAACAAAGTCCA 1933
1923 GCAAGAGATTTGATGAGATGCTGCTTACCTTACCTTAAAGAGCTTGAAGAGAGGAGGAGG 1982
1934 GCAAGAGATTTGATGAGATGCTGCTTACCTTACCTTAAAGAGCTTGAAGAGAGGAGGAGG 1993
1983 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2042
1994 ACTAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2053
2043 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2102
2054 ACTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2113
2103 ACAGCAACTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2162
2114 GCAAGCAACTCCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2173
2163 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2222
2174 ACAGCAACTCCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2233
2223 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2279
2234 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2293
2280 CCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2339
2294 CCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2353
2340 AGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2399
2354 AGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2413
2400 TAAAGTGGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2459
2414 TAAAGTGGCCACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2473
2460 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2519
2474 ACAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2533
2520 CGATGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2564
2534 TGAAGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2593
2565 AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2624
2594 AAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2653
2625 CCTGCAAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2684
2654 CCTGCAAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2713
2685 GACGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2744
2714 GACGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2773
2745 GACGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2804
2774 GACGAGATCACTATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2833
2805 GTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2864
2834 GTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2893
2865 GCTTCTCTGAGCGGTACTAGTGTGGCAGGTTACGGCAGAGATCAGGCTTATCCCGAGC 2924

Db 2894 GGCCTCTCAACGCTTCAAGTCCGCGCAAGTTAAAGCAGAGGCTCCGCTTTACTCCAGC 2953
QY 2925 CACAGCACTGGCTCCTCCCATCTCCGCTCGGGCCAGGGTGAAGAAAGGTGAAGAGGCT 2984
Db 2954 CACGGCACTGGCTCCTCCGCTCCTGCTGTAAGCCAGGGTGAAGAAAGGTGAAGAGGCT 3013
QY 2985 ACNAGCGCAAGCCCTGATCCCTGAGAGCCAAAGAAAGCAACCACTTAAATTTTAA 3044
Db 3014 ACAAGCTCAAGCCCTATCTCTTGGAGAGCCAAAGAAAGCAACCACTTAAATTTTAA 3073
QY 3045 AAGTACGTATCATCCGTTCTGGAACAGCAAGCATGTGTGTTTGGAGAGTTCAAG 3104
Db 3074 AATATGTATCATCCGTTCTGGAACAGCAAGCATGTGTGTTTGGAGAGTTCAAG 3133
QY 3105 TCAGAGGGTGTGTTCCCAAGCTTACGTGAACCTATTTCCAGGGCCGCTAGGAATC 3164
Db 3134 TCAGAGGGTGTGTTCCCAAGCTTACGTGAACCTATTTCCAGGGCCGCTAGGAATC 3193
QY 3165 CACNACATCATCTGCTGCTTCAAGTCCGCTGCTAGCTTAAAGAGAGTGGCTCC 3224
Db 3194 TACNACATCATCTGCTTCTTCAAGAGTCCGCTAGCTTAAAGAGAGTGGCTCC 3253
QY 3225 GGCCTCAAGCCAGCCATTCCTCCGAGAGAGTTTATTCATGTACATACGAGATTC 3284
Db 3254 AGCAGCCAGCCAGCTGTTCCGAGAGAGTTTATTCATGTACATACGAGATTC 3313
QY 3285 TAGAGAGAGATTTAACTTTTCAAGAGAGATGTATGTGTATCCAAAGAAATG 3344
Db 3314 TAGAGAGAGATTTAACTTTTCAAGAGAGATGTATGTGTATCCAAAGAAATG 3373
QY 3345 TGACTGTGAGACGGAAACGGTGGGCGACAAAGTCCGAGTCTTCCCTTCACTATGAG 3404
Db 3374 TGACTGTGAGACGGAAACGGTGGGCGACAAAGTCTTCCCTTCACTATGAG 3433
QY 3405 GCTTAAAGATTCAGAGGCTCTGGAACCTGTGGGAAACAGGAGTTTAAAGAAAC 3464
Db 3434 GCTTAAAGATTCAGAGGCTCTGGAACCTGTGGGAAACAGGAGTTTAAAGAAAC 3493
QY 3465 TGAATATTCAGAGGATTTATGCTTCTCAAGCTGTCTACTGTCTCCGAAACACTCC 3524
Db 3494 TGAATATTCAGAGGATTTATGCTTCTCAAGCTGTCTACTGTCTCCGAAACACTCC 3553
QY 3525 TCTTGGGAGCTGATCTGATCCGAGAAAGCAACGAGTGTGTGGGAGAGAGACT 3584
Db 3554 CCTGTGAGCTGATTTTATTCGAGAAAGCAACGAGTGTGTGGGAGAGAGACT 3613
QY 3585 GCAAGCTGAGGAGAAAGGCGCAGATAGGTTGTTCCAGCAATTAATGTCAACTCT 3644
Db 3614 GCAAGCACTGGGAGAAAGGCGCAGATAGGTTGTTCCAGCAATTAATGTCAACTCT 3673
QY 3645 AAGCCCGGAGAAAGCAAGAAATCCCAACTGAGCTAACAGCCGAGTGCAGCCAG 3704
Db 3674 AAGCCCTGGAGCAAGAAATCACTCCACAGGCCACTTAACTCAACGACTTACCGC 3733
QY 3705 AGTGTCCAGGTGATCGGATGTAGATTACACCGCCAGAAAGATGAGAACTAGCCT 3764
Db 3734 AGTGTCCAGGTGATCGGATGTAGATTACACCGCCAGAAAGATGAGAACTAGCCT 3793
QY 3765 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGGAAGAGA 3824
Db 3794 CAGCAAGGCGCAGATCATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGGAAGAGA 3853
QY 3825 AGTCACTGGGCAAGTTGGGCTCTTCCCATTCATTTATGAAGCTACCAAGACATGGA 3884
Db 3854 AGTCACTGGGCAAGTTGGGCTCTTCCCATTCATTTATGAAGCTACCAAGACATGGA 3913
QY 3885 CCCACCCAGCAATGATCATATATGTGTCTATCCCGCTCAGGCTTGAAGCTCCCAA 3944
Db 3914 CCCACCCAGCAATGATCATATATGTGTCTATCCCGCTCAGGCTTGAAGCTCCCAA 3973
QY 3945 GAGACCCACTATTCATATCACTGCCAGAGGATGATGGAGATCAGCTTATCATG 4004

Db 3974 GAGACCCACTATTCATATCACTGCCAGAGGATGATGGAGATCAGCTTATCATG 4033
QY 4005 TGACTTGCAGATGATCACTTACTGCTTCTGATGAGAAATCTCTGCAAGCAGTT 4064
Db 4034 TGACTTGCAGATGATCACTTACTGCTTCTGATGAGAAATCTCTGCAAGCAGTT 4093
QY 4065 ACCCTATTTGACCTTGTGATGATGATGAATGTCTGAGTCACTGCGGAGAGAGAG 4124
Db 4094 ACCCTATTTGACCTTGTGATGATGATGATGAATGTCTGAGTCACTGCGGAGAGAG 4153
QY 4125 AAGC--AAATTCAGAACTGCAAGGGTGTGGGCTTTTGGGCTTCTAGTCACT 4182
Db 4154 GAGCAAAATTCAGAAATCAGACAGGGTGTGGGCTTTTGGGCTTCTAGTCACT 4213
QY 4183 AAGTCAAGCCG--CCCGCTTCAACAGGGGCTTCAATAGTTTAAATTTTAA 4241
Db 4214 AATATGACTTTCCTCCCACTTTCAGACAGGTGTTCATAGTTTAAATTTTAA 4273
QY 4242 TGTATTTTACCTTTTAAATCTCAATCAATCTCTTGGCTTATTTGGTT 4301
Db 4274 TATATTTTACCTTTTAAATCAAAATTAATTAATCACTTCTTGTATTTGGTT 4333
QY 4302 TACAAACACCCACTATCAGAGAGTCTGTGCGAGCATTAATGTCTTCCG 4361
Db 4334 TCAAAACACCCACTATCAGAGAGTCTGTGCGAGCATTAATGTCTTCCG 4388
QY 4362 CGTACCTTAACTGAGAGCTGTCTGATCTTGGCGTTGTCTCACTTCCCAACT 4421
Db 4389 ATGTCAATTAATCTGAGACTTATGATTTTTCATTTTGTCTCAGGTTCACCACT 4448
QY 4422 TGTATGTTTGGGCTGTCCC---TGCGTAGAGACAGAGGATGGGTGATCT 4477
Db 4449 TGTGCACTTGGGCTTTCCTTCCCTTACATGAAAGTGCAGAGACTTCAATCTCT 4508
QY 4478 TTTGAATATGTATGTAGACTGAGCTGATGAGAGGGTTATGCTGTCTGTACC 4537
Db 4509 TTTAAAGAGTATA--GAATGAGCCCAATTAAAGCAGAGTGTGTGCTGTGTGT 4567
QY 4538 ATCAGCTGATC--TGTGCGCATGTCATCTGTACCGAAGATGATCT--CTTCTCAT 4595
Db 4568 ATCAGCTGATCCTTGTGAGCATGTATATCATCTGTATGAAGAAATGATCTTCCA 4627
QY 4596 GGTAAACCCACACCGGTATGAGTCTCTCACTACT--GGATCATTTTCTTTGAC 4653
Db 4628 TGCAGAGCTATTTACTTGTGATGATGCTTATATATGATTTATTTTGGAC 4687
QY 4654 AGTGACTTGTAGCAGCTGAGAGAC--CCATGTTTCCGTTGGTCTCAGATGACT 4712
Db 4688 AGTGACTTGTAGCAGCTGAGAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 4747
QY 4713 AGTGTGCCGCTGT 4772
Db 4748 GGTATGTTGT 4807
QY 4773 TAAAGCGCAATGCGCCAGGGTTACATATCATGATCCACCGTCT---TTAGTCTC 4827
Db 4808 GTAGACAC--ACCATGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4866
QY 4828 TGTACGTGAG--TTTATTCAGTTGCTTTTATGAA---TATCTGAAAGATAT 4881
Db 4867 TGTATCATGATTTTATTCAGTTACTTTTATGAAATGACCTATTTTGAAGATAT 4926
QY 4882 CTTCTTGCAGAAAGATGTATGAGTCTCTCTGCAATTAATTTCCAGTGTATCAT 4941
Db 4927 TTTCTTGCAGAAAGATGTATGAGTCTCTCTGCAATTAATTTTCCAGTGTATCAT 4985
QY 4942 TTTTAACTAGCTGTGGGGGTGTCTACAGATTAATGAAGTGGGCTCCGCTCCG 5001
Db 4986 TTTTAACTAGCTGT--GGAATTTCTACAGATTAATGAAGTGGGCTCCGCTCCG 5044
QY 5002 TGTGTATTAATCTTGT 5058
Db 5045 TGTGTGT--AGATATGT 5103

Db 3119 GTGAACCTCATTTTCAGGGCCCATAGAAGTCTACAGAGATGATTTGTTCTTCAAG 3178
 Qy 3193 AGTCTGTAGTCTTAAAGAGTGGCTTCCCGGCGCCCAAGCCATTCGCCGAGAA 3252
 Db 3179 AGTCTGTAGTCTTAAAGAGTGGCTTCCCGGCGCCCAAGCCATTCGCCGAGAA 3238
 Qy 3253 GAGTTATTTGCATGTACATACAGAGATTTCTGAGCAAGAGATTTAACTTTGAGCA 3312
 Db 3229 GAATTTATTTGCATGTACATACAGAGATTTCTGAGCAAGAGATTTAACTTTGAGCA 3298
 Qy 3313 GGGGATGTATTTGTTTACCAAGAAAGTGTGACTGCTGAGCGGAAACGCTGGCGAC 3372
 Db 3299 GGGGATGTATTTGTTTACCAAGAAAGTGTGACTGCTGAGCGGAAACGCTGGCGAC 3358
 Qy 3373 AAGTCCGAGTCTTCCCTTCAATGTAGAGCTTTAAAGATTCAGAGGCTTGGAACT 3432
 Db 3359 AAGGCGGAGTCTTCCCTTCAATGTAGAGCTTTAAAGATTCAGAGGCTTGGAACT 3418
 Qy 3433 GCTGGGAAAAACAGGAGTTTGAAGAAAAAACTGAAATTTGCCAGTTATTGCTCTAC 3492
 Db 3419 GCTGGGAAAAACAGGAGTTTGAAGAAAAAACTGAAATTTGCCAGTTATTGCTCTAC 3478
 Qy 3493 GCTGTACTGTCTCCGAACAACCTGCTGCTGAGGAGCTGATTTCTGATCCGAAA 3552
 Db 3479 ACCGCGACCGGCGCCGAGCAGCTCACTCGCGCCCTGGTCACTGATTTGATCCGAAA 3538
 Qy 3553 AAGAACCAAGTGTATGTGGGAAGAGAACTGCAAGCTCCAGGAAAAAGGCCAGTA 3612
 Db 3539 AAGAACCAAGTGTATGTGGGAAGAGAACTGCAAGCTCCAGGAAAAAGGCCAGTA 3598
 Qy 3613 GGGTGTCTTCCAGCAATTTATGTCAAACTTCTAAGCCCGGAAACAAGAAAAATCA 3672
 Db 3599 GGGTGTCTTCCAGCAATTTATGTCAAACTTCTAAGCCCGGAAACAAGAAAAATCA 3658
 Qy 3673 ACTGAGTACCCAGAACCGCAGTGCAGCAGAGTGTGCGCAGTGTATGGATGTAGAT 3732
 Db 3659 ACAGAGCCACCTTAAGTCAACAGCATTAGCGGAGTGTGCGCAGTGTATGGATGTAGAT 3718
 Qy 3733 TACACCGGCCAGAAAGATGACGAACTAGCTTACCAAAAGGCCAATATCAAGCTCTC 3792
 Db 3719 TACACCGGCCAGAAAGATGACGAACTAGCTTACCAAAAGGCCAATATCAAGCTCTC 3778
 Qy 3793 AACAGAGAGAACCGCAGTGTGGAAGAGAAAGTCAAGTGGGCAAGTTGGCTTTCCCA 3852
 Db 3779 AACAGAGAGAACCGTGAAGTGTGGAAGAGAAAGTCAAGTGGGCTTTCCCA 3838
 Qy 3853 TCCAAATATGTAAAGCTGACACAGACATGAGCCCGCAGCAGCAATGATCATATGTTGT 3912
 Db 3839 TCCAAATATGTAAAGCTGACACAGACATGAGCCCGCAGCAGCAATGATCATATGTTGT 3898
 Qy 3913 CCATCCCCCTCAGAGCTTTGAAATCTCTAAAGAACCCCATATCCCATATCACTGCCCA 3972
 Db 3899 CCATCCCCCTCAGAGCTTTGAAATCTCTAAAGAACCCCATATCCCATATCACTGCCCA 3921
 Qy 3973 GAGGATGTATGGAGATGACAGCTTATCATGTAGTCTGACAGCATGATCACTACTGCT 4032
 Db 3922 ----- 3921
 Qy 4033 TCTGATGAGAAAGTCACTGACGAGAGAGTTTACTCATTTGACCTTAGTGCATGTAT 4092
 Db 3922 ----- 3921
 Qy 4093 CGAATATGTAGTCACTGCTGCGAGAGCAGAACCAATTGCAAACTGACAGAGGTGG 4152
 Db 3922 ----- 3921
 Qy 4153 TGGGTCTTTTGGGCTTTCTCTAGTCACTGACAGACTGAC-CGGCCCGGCTTCAACAGGAGC 4211
 Db 3922 ---GTCCTTTTGGGCTTTCTCTAGTCACTGACAGACTGACCTTTCCCGCAGAGT 3978
 Qy 4212 GCTTCAATAGTATTTAAATATTTTAAATGTATTTTAACTTTTAAATATAAATCTC 4271
 Db 3979 GCTTCAATAGTATTTTAAATATTTTAAATATATATTTTAACTTTTAAATATAAAT 4038

Qy 4272 AATCAATTACTCTTTGGCTATTTTGGTTTACAAAAACACCACTAATCAAGAGTGGCT 4331
 Db 4039 AATTAATAGTCTTTGCTATTTTGGTTTACAAAAACACCACTAATCAAGAGTGGCT 4095
 Qy 4332 GTCGCGGACGATTTAAATGCTGTTCGCGGCGTACCGTAACTGAGAGCTTGTCTACT 4391
 Db 4096 --CTGCAATGCTATTTAAATTTGTTCCAAATGTCATTAATCTGAGAGCTTATGATTT 4153
 Qy 4392 TTGCGCTTTTGTCCAGTGTTCCTCAACACATGTTGTATGTTTGGGCTGTTCCTC---TGCC 4447
 Db 4154 TTTCAATTTTGTCCAGTGTTCCTCAACATTAATGTCAGTGTGGGCTTTTCCCTTACCA 4213
 Qy 4448 GTAGAGCAGAGAGATGGGTGATACCTGTTTGAATATGATATAGTACGTAGAGCTGA 4507
 Db 4214 TAGAATGTCAGAGAGATTCAGTATCTGTTTAAAGAGTATGAAATGAGCCCAATTA 4273
 Qy 4508 CTATGAAAGGGGTATGCTGTCTGTGACATCAAGTGTACC-TGTGCGCATGTACCAT 4566
 Db 4274 AGC-GAAGT 4332
 Qy 4567 CTGTAACGAGAGATGACT-CTTCTCTCATGCTTAAACCAACCGTGTACAGTGTCT 4625
 Db 4333 ATCTGTACATMAAGAAATGATGTTCTTCCATGCGCAAGCTATTAACCTGTACAGTGTCT 4392
 Qy 4626 CATCTACT--GCATTCATTTTACTTGTGACAGTACCTGTGAGCCCTGAGAGAGCACC 4683
 Db 4393 AATCATATGCTATTTAATTTTATTTTGTGACAGTACCTTGTAGCAGCATGAGAAAGCACT 4452
 Qy 4684 CA-TGTTTCCGTTTGTCTCAGATGTACTAGTGTGCGCGGTGTTTGTATTTTTC 4742
 Db 4453 CTGTGTTTGT 4512
 Qy 4743 AATCTGCAATCTTTCACACATTAATAGTAAGCCCACTGCCAGGCTTATGAT 4802
 Db 4513 AATTTTGGTGTTCATATGATTAATATGATTAATATGATTAATATGATTAATATGAT 4571
 Qy 4803 CATCAGTACCCAGGCTC-----TTAGTCTGTGACGGAAG-TTATATTCAGTGTCTT 4856
 Db 4572 CAAGCATTCACAGTCTCTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4631
 Qy 4857 TTATGAA-----TATCTGAACAAGTATCTTCTTGAACAAGAAATGATTAAGATC 4911
 Db 4632 TCATGAGATGACCTATTTTGAACAAGTATTTCTTGAACAAGAAATGATTAAGATC 4691
 Qy 4912 TCCCTGCAATTAATTTCCAGTGTTCATATTTTAACTAGACTGTGGGCTGTACAG 4971
 Db 4692 TCCCTGCAATTAATTTT-CCATGTTTACATTTTAACTAGACTGT-GGAAATTTCTACAG 4749
 Qy 4972 ATTAATATGAAGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5031
 Db 4750 ATTAATATGAAGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4808
 Qy 5032 GTGTGT---CCTTAGATATTTAGTGAAGTGGGAGAGATTT 5071
 Db 4809 CTGTTTGTCTTTAAACCTAGTTGAAGCTCTCAATTAATAT 4851

RESULT 7
 AA234570
 ID AA234570 standard; cDNA; 5199 BP.
 XX AA234570;
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA.
 XX
 DE SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; chromocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.

KW ss.
 XX Homo sapiens.
 OS
 XX
 XX
 FT Key Location/Qualifiers
 CDS 208..3642
 FT /*tag= a
 XX
 XX MO953062-A2.
 XX
 XX 21-OCT-1999.
 XX
 XX 16-APR-1999; 99WO-US08371.
 XX
 XX 16-APR-1998; 98US-0082007.
 XX
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 XX Korenberg JR, Chen X;
 XX
 XX MPI: 1999-633829/54.
 DR P-PSDB; AAY32154.
 XX
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX
 XX Claim 2, Fig 5, 99pp, English.
 PS
 XX
 XX This is the nucleotide sequence of full-length cDNA corresponding
 CC to a novel human SH3 gene, termed the SH3D1A gene, that contributes
 CC to the development of platelets and the pathogenesis of leukaemias,
 CC both in general and in particular those involving the small
 CC megakaryocytic lineage. The SH3D1A gene maps to the small
 CC candidate region for low platelets on chromosome 21. Sequencing
 CC of 5 different sizes of cDNA clone (see AAY34570-74) suggests that
 CC at least 3 isoforms exist. The invention provides methods for the
 CC diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for: suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder.
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 CC
 XX
 XX Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;
 SQ
 Query Match 60.6%; Score 3080.2; DB 20; Length 5199;
 Best Local Similarity 79.8%; Pred. No. 0;
 Matches 4030; Conservative 0; Mismatches 753; Indels 268; Gaps 23;

QY 235 CAGCAAGTGAAGTATAGAAACCATGGCTCAGTTTCCACACCTTTGGTGTAGCTG 294
 |||||
 Db 184 TAGCAAGGTAAAGTAAAGTAAAGCAATGCTGCTCAGTTTCCACACCTTTGGTGTAGCTG 243
 QY 295 GATGTCTGGGCACTAACTGTGAGAGAAAGGCGCAACATGACGAGGATCTTACCTG 354
 |||||
 Db 244 GATATCTGGGCACTAACTGTGAGAGAAAGGCGCAACATGATCAGAGTCTTCAATGTTTA 303
 QY 355 AAGCCGATAGCGGAGATTATTTACTGTGATCAAGCAGAGAACTTTTTCATCTGGG 414
 |||||
 Db 304 AAGCCGATATCTGGATTTCACTGTGATCAAGCAGAGAACTTTTTCATCTGGG 363
 QY 415 TTACTCAGCCTGCTTTAGCACAATAATATGGGCGCTGAGGAGCATGATAAGTGAAG 474
 |||||
 Db 364 TTACTCAGCCTGCTTTAGCACAATAATATGGGCGCTGAGCAGTATGATATGATGAAG 423
 QY 475 ATGGATCAAGTGAATTTTTCATAGCAGTAAAGCTTATCAAACTGAAGTACAAAGATAT 534
 |||||
 Db 424 ATGGATCAAGTGAAGTTTTTCATAGCAGTAAAGCTTATCAAACTGAAGTACAAAGATAT 483
 QY 535 CAGCTCCCTCCACACTTCCCTCTGATGAAACAGCAACAGTGGCTATTTCCAGTGA 594
 |||||
 Db 484 CAGCTACCTCTGCACTTCCCTCTGATGAAACAGCAACAGTGGCTATTTCCAGTGA 543
 QY 595 CCAGCATTTGGTATGAGAGGATGCTAGCATGCAACGCTACAGCTGTGCTGCTG 654
 |||||
 Db 544 CCAGCATTTGGTATGAGAGGATGCTAGCATGCAACGCTACAGCTGTGCTGCTG 603
 QY 655 CCAATGGGCTCATTCCAGTTGTAATGTCACACCTTATGATCTTCTGCTCCCA 714
 |||||
 Db 604 CCAATGGGCTCATTCCAGTTGTAATGTCACACCTTATGATCTTCTGCTCCCA 663
 QY 715 GCAAGCATGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 774
 |||||
 Db 664 GCAAGCATGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723
 QY 775 GCGCATCTGCAAGCCACATGGCCAAAGATTTCTCTTGAAGAGATCTGCTCAGGCTCA 834
 |||||
 Db 724 GCTCATCTGCAAGCCACATGGCCAAAGATTTCTCTTGAAGAGATCTGCTCAGGCTCA 783
 QY 835 CAATTAACTAATAGTTAAGAAAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTG 894
 |||||
 Db 784 CAACTAACTAATAGTTAAGAAAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTG 843
 QY 895 GCAAGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 954
 |||||
 Db 844 GCAAGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 903
 QY 955 GCAAAATCTAATAGTTAAGAAAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTG 1014
 |||||
 Db 904 GCAAAATCTAATAGTTAAGAAAGGACATGATGCTGCTGCTGCTGCTGCTGCTGCTG 963
 QY 1015 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTCGACATTTGATCAAGATGA 1074
 |||||
 Db 964 AGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTCGACATTTGATCAAGATGA 1023
 QY 1075 AAATCTACTGCAAGAAATTTATCTAGCTATGCACTAATGATGTTCCATGTCTGT 1134
 |||||
 Db 1024 AAATCTACTGCAAGAAATTTATCTAGCTATGCACTAATGATGTTCCATGTCTGT 1083
 QY 1135 CAGCCATGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1194
 |||||
 Db 1084 CAGCCATGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1143
 QY 1195 TCCGCAATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1254
 |||||
 Db 1144 TCCGCAATGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1203
 QY 1255 CCGTCTGCAAGAGATGAGCAAGC---CAGAAAGAAATGCTGCTGCTGCTGCTGCTGCTG 1311
 |||||
 Db 1204 CCAAGTTTATGAGAGATGAGCAAGCAATTTAGAAAGAAATTTAGCTGTTGAAGAT 1263
 |||||
 QY 1312 AAGAAAGCGGAGAACTTGAAGAGGAGGATGAGAGCTGGAAGAGCGCCCAAGGCTC 1371

1264 AAGAAAGCGGAGAACTTTGAAGTGGCAACTGGAACTGGAGAAACGAAGGAACTCTC 1323
1372 TTGAGCAGCAGCGCAAGAGAGAGAGCGGTTGGCTCAGCTGGAGCGCGGAGAGAG 1431
1324 CTGGAAACGAGCGCAGAGAGAGAGCGCTGGCCAGCTGGAGCGGCGGAGAGAG 1383
1432 AGGAAAGCGGAGAGAGAGAGAGAGAGCGCAAGCGGAGCTGGAGCTGGAGAGAG 1491
1384 AGAAGAGAGCGTGAAGCGCAGAGAGAGAGCGCAAAAGACAACTGAACTGGAGAG 1443
1492 CTGGAGAGAGAGCGGAGCTGGAGCGGAGCGCAGAGAGAGAGAGAGAGAGAGTCCAG 1551
1444 CTGGAGAGAGAGCGGAGCTGGAACGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1503
1552 AGCGCGAGCGCGCAAAACGGGAACTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
1504 AGCGAGAGAGCGTGGCAAAACGGGAACTGGAGAGAGAGAGAGAGAGAGAGAGAG 1563
1612 CGGAGAGAGAGAGCTGGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
1564 CGAGAGAGAGAGAGCTGGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1623
1672 GCAAGAGAGAGAGAGCTGGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
1624 GCAAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1683
1732 GAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAGAG 1791
1684 GAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAGAG 1743
1792 AGGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1851
1744 AGGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1803
1852 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1911
1804 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1863
1912 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1971
1864 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1923
1972 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2031
1924 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 1983
2032 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2091
1984 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2043
2092 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2151
2044 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2103
2152 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2211
2104 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2163
2212 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2268
2164 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2223
2268 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2328
2224 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2283
2328 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2388
2284 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2343
2388 GAGAGAGAGAGAGAGCTTGGAAATTTGAAATGAGAGCTTGAATGAGAGAGAGAG 2448

2344 CAAGAACAGAGTAAAGCAGAGTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2403
2449 ACATTTCTGAG 2508
2404 ACATTTCTGAG 2463
2509 TCCAG 2568
2464 TCCAG 2523
2569 CAG 2628
2524 CAG 2583
2629 GAG 2688
2584 GAG 2643
2689 GAG 2748
2644 GAG 2703
2749 GAG 2808
2704 GAG 2763
2809 GAG 2868
2764 GAG 2823
2869 GAG 2928
2824 GAG 2883
2929 GAG 2988
2884 GAG 2943
2989 GAG 3048
2944 GAG 3003
3049 GAG 3108
3004 GAG 3063
3109 GAG 3168
3064 GAG 3123
3169 GAG 3228
3124 GAG 3183
3229 GAG 3288
3184 GAG 3207
3289 GAG 3348
3208 GAG 3207
3349 GAG 3408
3208 GAG 3207
3409 GAG 3468
3208 GAG 3207
3469 GAG 3528
3211 GAG 3270


```

QY 3529 GGGCAGCTATTCGATCCGGAAGAAAGCCAGTGTGTGGAGAGAACTGCAA 3588
DB 3271 GGTCACTGATTTTATCCGAAAAAGAACCCAGGTGTGTGGAGAGAGCTGCAA 3330
QY 3589 GCTCGAGGAAAAAGCCAGATAGGGTGTTCAGCAAAATTATGTCAAATCTTAAAC 3648
DB 3331 GCACGTGGGAAAAAGCCAGATAGGGTGTTCAGCAAAATTATGTAAAGCTTCTAAAG 3390
QY 3649 CCGGGAACAAGCAAAATCAACCCCACTGAGCTACCAAGACCGCATGTGACGACAGTGTG 3708
DB 3391 CCGGGAACAAGCAAAATCACTCCAGAGCCCACTGAGCTACCAAGATTAAGGAGAGTGTG 3450
QY 3709 TGGCAGGTATCGGATGTACATTCACCGCCAGAGAGATGAGCAATAGGCTTCAGG 3768
DB 3451 TGGCAGGTATGTGGATGTACATTCACCGCCAGAGATGAGCAATAGGCTTCAGG 3510
QY 3769 AAAGGCAATCATCAACGCTCTCAACAGAGAGACCCGAGCTGTGTGAAAGAGAGATGT 3828
DB 3511 AAAGGCAATCATCAACGCTCTCAACAGAGAGACCCGAGCTGTGTGAAAGAGAGATGT 3570
QY 3829 AGTGGCAATGTGGGCTTTTCCATTCATTAATGTAAAGCTGACACAGCATGTGACCC 3888
DB 3571 AATGACAAAGTGGGCTTTTCCATTCATTAATGTAAAGCTGACACAGCATGTGACCC 3630
QY 3889 AGCAGCAATGAATCATATGTGTGTCCATCCCTCAGGCTTGAAGTCTCAAGAGAG 3948
DB 3631 AGCAGCAATGAATCATATGTGTGTCCATCCCTCAGGCTTGAAGTCTCAAGAGAG 3690
QY 3949 CCCATATCCCATATCATCTGCCAGAGAGATGATGTGGAGATGTGAGCTGTGATGTGAC 4008
DB 3691 CCCATATCCCATATCATCTGCCAGAGAGATGATGTGGAGATGTGAGCTGTGATGTGAC 3750
QY 4009 TTGCAGCATGATCACTTACCTTCTGTGATGAGAGAAATCATCTGACAGACGATTTA 4068
DB 3751 TTGCAGCATGATCACTTACCTTCTGTGATGAGAGAAATCATCTGACAGACGATTTA 3810
QY 4069 CATTTAGCTTGTGATGATGATGCAAAATGTCTGATGTCAGGCTGACAGAGCAGAG 4128
DB 3811 CATTTAGCTTGTGATGATGATGCAAAATGTGATGATGATGATGATGATGATGATGATG 3870
QY 4129 --AATTTGAGAACTGACAGAGGTGTGGTCTTTTGTGGGCTTCTTGTGCTGCTGAC 4186
DB 3871 AAAAATTTACAAAACACACAGGATAGTGGTCTTTTGTGGGCTTCTTGTGCTGCTGAC 3930
QY 4187 TGAC--CGGCCCCGCTTCAACAGCGGCTTCAATAGTTTAAAGTATTTTAAATGTG 4245
DB 3931 TGACCTTTCCCACTTGTGACAGGTGCTTCAATAGTTTAAATTTTAAATTTTAAATATA 3990
QY 4246 TATTTAGCTTGTGATGATGATGCAATCAATCAATCTTGTGCTGATTTGCTTTTACA 4305
DB 3991 TATTTAGCTTGTGATGATGCAATCAATCAATCTTGTGCTGATTTGCTTTTACA 4050
QY 4306 AAAACACCCATATCAAGAGGTGCTGTGTGCGAGATTAATCTGTCCGCGCGTA 4365
DB 4051 AAAAGACCCATATCAAGAGATG-----CTGCAATGTGATTAATTTGTTCCAAATGT 4105
QY 4366 CCGTAACTGAGAGCTTGTGATGCTTTGTGCGTTTGTCCAGTGTCCCAACCATTTG-T 4424
DB 4106 CCAATTAATCTGAGATGATGATGATTTTCAATTTTGTCCAGTGTCAACATTAATGTCT 4165
QY 4425 GTAGTTTGGGGCTGTCCC-----TGCGTGAAGACAGAGAGATGGGTGTGCTGCTTTT 4480
DB 4166 GCGATTTGGGGCTTTTCCCTTACATTAAGATGTGAGAGATGAGATGATGATGCTGTGTTT 4225
QY 4481 GAAATGTGTATGTATGATGAGCTGATGATGAGAGAGGATGATGATGCTGTGTGACATC 4540
DB 4226 AAAAGATGTATGATGATGAGCTGATGATGAGAGATGATGATGCTGTGTGATGATC 4284
QY 4541 AGCTGTACC-TGTGCGGATGTACATCTGTGATGAGAGAGATGATGCT-CTTCTGCAATGCG 4598
DB 4285 AGCTGTACCTTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4344

```

```

QY 4599 TAAACCCACACCGGTGATGATGCTGTCTCATC-TACGTACATTCATTTTACTTT--GCACAG 4655
DB 4345 CAAGCTATTAATCTTGTGATGATGCTGTCTCATC-TACGTACATTCATTTTACTTTTTCACAG 4404
QY 4656 TGACCTTGTAGCCACCTGAGAGAGAC--CATGTTTCCGTTTGTGTCTGATGATGATGATG 4714
DB 4405 TGACCTTGTAGCCACCTGAGAGAGACCTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4464
QY 4715 TTGTGCGGCTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4774
DB 4465 TTGTGCGGCTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4524
QY 4775 AGACGCCATCTGCCAGGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4829
DB 4525 AGACGCCATCTGCCAGGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4583
QY 4830 TTACGTGAAG-TTATTCAGTGTGCTTTTATGAA-----TATCTGAAACAATGATCT 4883
DB 4584 TTACATGAAGTTTATTCAGTGTGCTTTTATGAA-----TATCTGAAACAATGATCT 4643
QY 4884 TCTTGACAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4943
DB 4644 TCTTGACAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4702
QY 4944 TTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5003
DB 4703 TTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4762
QY 5004 TGTGTAACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5060
DB 4763 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4821
QY 5061 GGAAGGAATT 5071
DB 4822 TCAATTAATAAT 4832

RESULT 8
AAS84763
ID AAS84763 standard; cDNA; 7435 BP.
XX
AC AAS84763;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20567.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PP 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
P-PSDB; ABG20576.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

```


OY	1734	AGGAAAACCTTCAGATATCAGGTGGTGCATCGGCCAACCCAGAGGCGAAGAATTGAGAC	1793
Dd	1751	AGGGAACCTTCAAGATATCAGATGTCGATTGACCACCCAAAGGCGAAGAATTGAGAC	1810
OY	1794	GAAACAAGTCTAGAGACTAGAATTGCTG--AAATCACCCCATTTACACAGACAGT--GC	1849
Dd	1811	AAACAAATCTAGAGAGTGAGATTGCCGGAATATCAACCCCATCTACACATAATTACA	1870
OY	1850	AGGAATCTCAGCAA--TGCTTGAGACTTTATTCAGAGAAACAGATACTCAGTACC	1906
Dd	1871	TGGAATCTCAGCAAAATGCTTTGGAGAAGACTTAATCCAGAAAAACAGATATCTCATGTACC	1930
OY	1907	AGTTAAACAGSTCCAGAGAAACAGTTTGCAATAGAGACTCGCTTTACCTCAAAAG-A	1965
Dd	1931	AATTAAACAAGTTAGAGAAACAGTTTGCAACAGATTCCTGTTACACTTAAAGAA	1990
OY	1966	GCTTTGAGAGCAAAAGAGCTGGCCCCGAGCAGAGCTCCGAGAGCAGCTGAGCAGGTGAG	2025
Dd	1991	GCTTTAGAGCAAAAGAACTAGCTCGAGACACTTACAGAACCAACTGATGAATGGAG	2050
OY	2026	AGAGAGACCAAGTCAAGCTGCAGAGATTGATGTTTTCAACACACTGAAGAACTG	2085
Dd	2051	AAAGAAATAGATCAAACTACAGAGATTGATTTTCAATTAATCACTGAAGAACTA	2110
OY	2086	AGAGAGATCATAGCAAAACGAACTCCAGAAAGCAGAGTCCCTGAGAGCCGCGACTG	2145
Dd	2111	AGAGAAATATCAATATAGCAACATCTCAGAAAGCAAAAGTCCATGAGAGCTGAACGACTG	2170
OY	2146	AAGCAGAAAGCAGAGAGAGAGAGCTGGAGTTAGAGAAAGCAAAAGAAAGAGCGCTAG	2205
Dd	2171	AAACAGAAAGAACAGAACGAAAGATCATAGATTAGAAAACAAAAAAGAAAGCCCA	2230
OY	2206	AGACGAGTTCAAGAAAGGAGCAAGCAATGCTGAGCATGTGCACAGAG- --GAGCAG	2262
Dd	2231	AGACGAGCTCAGAAAGGAGCAAGCAGTGGCTGAGCATGTGCACAGAGAGAGAGAT	2290
OY	2263	CCAAGCCCCCGGAAAACCCCAAGAGAGACAGACTGAAGAGGAAAGACGTGTCAAGAG	2322
Dd	2291	CAGAACCCMAAGAAAATCTCAGAAAGGAAAAATCTGAAAAGGAGAGAGTGTCAAAAAG	2350
OY	2323	AAGGA-GGCGGAAAGAGAGCGCAAGCCGGAATGTGAACAAGACAGAGTGGCTTTTCCA	2381
Dd	2351	AAGATGCAAGAGAAAAAGGCAACAGAAAGCAACAGACAGCTGGGTCCGCTTTTCA	2410
OY	2382	TCCGATCAGAGCCAGCTAAGCTGGCCACCCAGCAGACCTCGTGTACACAGAGAAAG	2441
Dd	2411	TCAACACMAAACCAGCTTAGCCAGCTGTCAAGCACCTCGTGTCACTGACGAAAAAG	2470
OY	2442	CCCGCTTACATTTCTGCACAGAGAGTGTAAAAAGTGTATATTACGAGCGCTGTACC	2501
Dd	2471	TCCACTTACATTTCTGCACAGAGAAATGTAAAAAGTGTATATTACGCGCAGCTGTACC	2530
OY	2502	CTTTGAATCCAGAAAGTCAGATGAGATTCACATCCAGCAGAGATATATGTCAT-----	2555
Dd	2531	CTTTGAATCCAGAAAGTCAGATGAGATTCATATCCAGCAGAGACATATGTCAGTTAA	2590
OY	2556	-----GGTGGATGAAGACCCAGACTGAGAGAGCAGAGTGGCTTGGAGAGAGCTGAA	2606
Dd	2591	AGGGGAATGGGTGATGAAGAACCCAACTGAGAAACCCGGCTGGTGGAGAGAAATTAA	2650
OY	2607	AGGGAGACGGGATGCTTCCCTGCACAACTATGACAGAAAAAGATTCCAGAAAAATGAGTTCC	2666
Dd	2651	AGGAAAGACAGGGGTGTTCCCTGCACAACTATGACAGAAAAATCCAGAAAAATGAGTTCC	2710
OY	2667	CAGTCAGGCCAAACAGATGCCATGTGACATCTGGCCCCCTGCCAAACTGGCTCTGGC	2726
Dd	2711	CGCTCAATGAACACAGTACCTGATTCAACATCTGCCCTTCCCAACTGGCTTGGCG	2770
OY	2727	TGAGACCCCTCTCTTTGGCAGTGAACCTTTCTGAGGCCCTGCACAAACCCCACAACTG	2786
Dd	2771	TGAGACCCCGCCCTTTGGCAGTGAACCTTTCTGAGAGCCCTTGCAGACCCCTTAATTAATG	2830
OY	2787	GGCAGACTTCAGTTCCACGTGGCCAGCAGCTCAAAACGAGAAACGACAAACGAACTG	2846

Dp	2831	GGCCACTTCAGCTCCACGTG3CCCAACGACAGAAATGAAACCAAGAAAGGATTA	2890
Qy	2847	GGATTAAGTGGGCGGGCTCAG-CTTTCCTGACCGTAACTGATGCTGGCCAGTTACGGCAGA	2905
Dp	2891	GGATTCAGTGGCGACGCGAGGCCCTTCTCTCACCCTTCCAAAGTCCGCGCAATTTAAGCCAGA	2955
Qy	2906	GATCAGCTTTTACCCTCAGCCACAGCCACTGGCTCTCTCCCATCTCCCGCTCTGGCCAGG	2965
Dp	2951	GGTCGCGCTTAATCTCAGCCACGCGCACTGGCTCTCTCCCGCTCTCTGTGTAAGCCAGG	3010
Qy	2966	GTGAAAAGTGTGAAGGGGTACAAAGCCGACGCCCTGTATCCTTGGAGAGCC-AAAAAGAC	3022
Dp	3011	GTGAAAAGTGTGAAGGGGTACAAAGCCCTAATCTTGTGAGAGGCCAAAAAAGAC	3070
Qy	3025	AACCACTTAATTTTAAACAAAAGTGAAGCTATCACCGTTCTGGAACAGCAAGACATGTGG	3088
Dp	3071	AACCACTTAATTTTAAACAAAAGTGAAGCTATCACCGTTCTGGAACAGCAAGACATGTGG	3130
Qy	3085	TGGTTTGGAGAAGTTCAGAGTTCAGAGGGTGTGTTCCCAAGTCTTAAGTGAATCTCAT	3144
Dp	3131	TGGTTTGGAGAAGTTCAGAGTTCAGAGGGTGTGTTCCCAAGTCTTAAGTGAATCTCAT	3190
Qy	3145	TCAGGCGCCGTAAAGGAATCCACAGCAT-CGATACTGGCCCTACTGAAAGTCTGCTAG	3203
Dp	3191	TCAGGCGCCGTAAAGGAAGTCTTCAAGAGATGGGATTCGTGTTCTTCAGAGAGTCCGTAG	3250
Qy	3204	TCCTAAGAGATGGGCTTCCCGGCGCGACAGCCACATTCGCGGAGAA-GAGTTTATG	3262
Dp	3251	TCCTAAGAGATGGGCTTCCCGGCGCGACAGCCACATTCGTGCGAGAGAAATTTATG	3310
Qy	3263	CCATGTACATACGAGAGTTCGTAG--CAAGAGATTTAACTTT--CAGCAAGGGGAT	3318
Dp	3311	CCATTTACATTAAGAGAGTTCGAGGAGAGGAGATTTAACTTTTACGCAAGGGGAT	3370
Qy	3319	GTGATTTGTGTTACCAAGAAAGATGTGTACTGTGTGACGGGAAAC-GGTGGCGACAAATC	3377
Dp	3371	GTGATTTGTGTTACCAAGAAAGATGTGTGACTTGTGTGACAAAGAAAGTGGCGCAAGAGC	3430
Qy	3378	CGAGATCTTCCCTCTAATCTATGTGAGGCTTAAAGATTCAGAGGGCTCTGGAATCTGCG	3437
Dp	3431	CGAGATCTTCCCTCTAATCTATGTGAGGCTTAAAGATTCAGAGGGCTCTGGAATCTGCG	3490
Qy	3438	GAAGAACAGGAGTTTAGGAAAAAAACCTGAAATTG-CCGAGTATTGCTTCTTACGCTG	3496
Dp	3491	GAAGAACAGGAGTTTAGGAAAAAAACCTGAAATTGCTTCTTACGCTTCTTACGCTG	3550
Qy	3497	CTACTGTCCCGAACCACTCACCTGTGCTCTCTGGGAGCTGATTTCTGATCCGAAAAAGA	3556
Dp	3551	CCACCGGCGCCGAGCAGCTCACTCTCGCCCTCTGTGCTGATTTTATTCGAAAAAGA	3610
Qy	3557	A-CCCAAGTGTATGTGTGGAGAGAGAACTGCAAGCTGAGAGGAAAAAGCCGCAATATGGG	3615
Dp	3611	ACCCAGGTGTATGTGTGGAGAGAGAACTGCAAGCTGAGAGGAAAAAGCCGCAATATGGG	3670
Qy	3616	TGTTTCCAGCAAAATTAATGTCAAACTTCTAAGCCCGGAAAACAAGCAAAATCACCCCACT	3675
Dp	3671	TGTTTCCAGCAAAATTAATGTGAAGCTTCTAAGCCCTGGAGAGCAAAATCATCTCCAA	3730
Qy	3676	GAGGTACCCAGACCGCAGTGCAGCCAGAGTGTCCAGGTGATCGGAGTGTACGATTAAC	3735
Dp	3731	GAGGTACCCAGTACAGCATTAAGCGGAGTGTGCCAGGTGATTTGGAGTGTACGATTAAC	3790
Qy	3736	ACCGCCCAAGAACGATGACGAATCTGACCTTCAAGCA-AGCCCAATCATCAACGCTCTCA	3794
Dp	3791	ACCGCCCAAGAACGATGACGATGAGCTTCAACAAAGGGGCGAAGTATCAACGCTCTCA	3850
Qy	3795	CAAGAGAGACCCGAGCTGTGTGAAGAGAGAAAGTCACTGTGGCAAGTTGGGCTCTTCCATC	3854
Dp	3851	CAAGAGAGACCCCTGACTGTGTGAAGAGAGAAAGTCAATGTACAAAGTGGGCTCTTCCATC	3910
Qy	3855	CAATTAATGAAGCTGACCAAGCATGACCCCGAGCCAGCAATGAATCATATGTTTCC	3914

Db 3911 CAATTATGTAAGTGCACGACATGGAGCCAGCCAGGACGATGTTGACACTTAC 3970
Qy 3915 ATCCCCCTCAGGCTTGAAAGTCTTCAAGAGACCCCTATCCATATC 3964
Db 3971 ATCTTGGATATGTGACCCCACTGAAGAAAGCAAGATATCATC 4020

RESULT 9
AA163825
ID AA163825 standard; cDNA, 3466 BP.
XX
AC AA163825;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 33.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antiinflammatory; antidiabetic; anticonvulsant; antibacterial;
KW antiparasitic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
XX
OS Homo sapiens.
XX
FN WO20015308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241121.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-488781/53.

P-PSDB; AAM43519.

New isolated nucleic acids and polypeptides, useful for diagnosing, treating and/or preventing human diseases and disorders -

Claim 1, SEQ ID NO 33; 664bp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AA163803-AA164012) and
 CC the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The
 CC genes were isolated from a range of human tissues disclosed in the
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at: ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;

Query Match 43.1%; Score 2188.8; DB 22; Length 3466;

Best Local Similarity 84.7%; Pred. No. 0;

Matches 2525; Conservative 0; Mismatches 427; Indels 30; Gaps 5;

QY 7 GAGGAGAGTGGAGCGCGCGGAGGCGGAGCGAGCTTGGTCTCCGTAGTACGGCGCT 66
 Db 35 GAGAAAGTGGAGGCGCCCGAGGAGGAGCGTAGCTTGTGTCTCCGTAAGTACGGCGCT 94
 QY 67 CGCAAGGAGCATCCCGAGCGGCTCCGGAGCGCGCGGAGGAGCGAGCGAGCGCGCG 126
 Db 95 CCGAGGAAGATCCCGAGCGGCTCCGGAGCG-----GACAGAGAGCGCGCGCG 143
 QY 127 GGGATGTGTGGCGGCTGCGGACTGCGGCTTCTGCG-GCGGCGTGGCGGCTGCACTGA 185
 Db 144 GGGATGTGTGGCGGCTGCGGCTCCTGCTCTCCAGCGCGCGGTGAGCGGCACTGA 203

QY 186 TTTGTGTGAGGCGCGCGCGCGCACCGCCCGAGATGAGGCGTGTGATCAGCAAGTGA 245
 Db 204 TTTGTCCCTGGGCGGCGGAGCGCGCGCGCGCGGAGATGAGGCGTGTGATGAGGTA 263
 QY 246 ACGTAATAGAACCATAGCTCACTTTCCACACTTTTGGGTGTAGCTGTGTCTGGCG 305
 Db 264 AAGTAACAGAACCATAGCTCACTTTCCACACTTTTGGGTGTAGCTGTGTCTGGCG 323
 QY 306 CATTACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 365
 Db 324 CATTACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 383
 QY 366 GGGATTTATTAAGTGTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 425
 Db 384 TGGATTTATTAAGTGTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 443
 QY 426 TGTCTTAGACAAATATGAGGCGCTAGCGGAGCATGATATGATGAGGAGGAGGAG 485
 Db 444 TGTCTTAGACAGATATGAGGCGCTAGCGGAGCATGATATGATGAGGAGGAGGAG 503
 QY 486 GGAATTTTTCATAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 545
 Db 504 GGAATTTTTCATAGCATGATGATGATGATGATGATGATGATGATGATGATGATG 563
 QY 546 CACACTTCCCTGTCATGAAACAGCAACAGTGTGCTATTCAGTGTGACCGCATTTGG 605
 Db 564 TGCACCTTCCCTGTCATGAAACAGCAACAGTGTGCTATTCAGTGTGACCGCATTTGG 623
 QY 606 TATAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 665
 Db 624 TATAGAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 683
 QY 666 CATTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 725
 Db 684 CATTCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 743
 QY 726 TCCCTGAGTAAAGGAGGCGCTCCCGGTACAGCTGAGCTGCTGCTGCTGCTGCTG 785
 Db 744 CCCCCTGAGTAAAGGAGGCGCTCCCGGTACAGCTGAGCTGCTGCTGCTGCTGCTG 803
 QY 786 AGCCATGAGGCGCAAGAGTCTTCTTCCAGCATGATGATGATGATGATGATGATG 845
 Db 804 AGCCATGAGGCGCAAGAGTCTTCTTCCAGCATGATGATGATGATGATGATGATG 863
 QY 846 TAAATTAAGAAAGGAGCATATGATGATGATGATGATGATGATGATGATGATGATG 905
 Db 864 TAAATTAAGAAAGGAGCATATGATGATGATGATGATGATGATGATGATGATGATG 923
 QY 906 TGTGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 965
 Db 924 TGTGCTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 983
 QY 966 GAGTGAACACTTAAC-----AGGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 1013
 Db 984 GAGTGAACACTTAACAGGTTCTGTTAGTGTCCCAAGCAAGAACTATTTTATGACGTC 1043
 QY 1014 AAGTTTACCAAGGCTCAGTGTGCTTCAATATGATGATGATGATGATGATGATG 1073
 Db 1044 AAGTTTACCAAGGCTCAGTGTGCTTCAATATGATGATGATGATGATGATGATG 1103
 QY 1074 AAAATCTCAGAGAAATTAATTCATGATGATGATGATGATGATGATGATGATGATG 1133
 Db 1104 AAAATCTCAGAGAAATTAATTCATGATGATGATGATGATGATGATGATGATGATG 1163
 QY 1134 TCAGCACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1193
 Db 1164 CCAACCACTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1223
 QY 1194 CTCGCGAGTGGATGTCGTCATTAAGTCTTTCTTGTGATGATGATGATGATGATG 1253
 Db 1224 ATCTGGCAGTGGATGTCGTCATTAAGTCTTTCTTGTGATGATGATGATGATGATG 1283
 QY 1254 GCCGTGTCAGAGGATGAGCAAGC---CAGAGAGAAATCGCTGTGATGATGATGAG 1310

XX Claim 1; SEQ ID No 88; 837bp; English.

CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiodystonia, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis. Infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastroenteric disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infection. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 42.1%; Score 2140.8; DB 23; Length 3319;

Best Local Similarity 85.0%; Pred. No. 0;

Matches 2447; Conservative 0; Mismatches 412; Indels 19; Gaps 4;

QY 111 GGCAGCGGGGCGGGGAGTGTGCGGCGTGGCGGCTGGGCTTCTGCG-GGAGC 169
DB 10 GACAGAGAGCGCGGGGATGTGTGCGGCGCTGGCTTCTGCGCTCCAGCGCG 69
QY 170 GTGCGGCGTGCACGATTTGTGTGAAGGGCGCGCGCACCCCGCGAGATGAGCG 229
DB 70 GCGTAGCGGCACTGATTTGTCTCGGGGCGGCGAGCGGAGCCCGCGAGATGAGCG 129
QY 220 TCGATCAGCAGGTGAAGTAACTGATGACCATGCTGCTTCCACACCTTTCGTGTA 289
DB 130 TCGATTTACAGGTGAAGTAACTGATGACCATGCTGCTTCCACACCTTTCGTGTA 189
QY 290 GCGTGAAGTGTGCGGCGTAACTGTGAGAGAAAGGGCGAGATGACGAGATGCTTA 349
DB 190 GCGTGAAGTGTGCGGCGTAACTGTGAGAGAAAGGGCGAGATGACGAGATGCTTA 249
QY 350 GCGTGAAGTGTGCGGCGTAACTGTGAGAGAAAGGGCGAGATGACGAGATGCTTA 409
DB 250 GTTTAAAGCGCAATATCTGATTCATTTACGATGATCAAGTAACTTTTTCAT 309
QY 410 CTGGGTACTCTGAGCTGCTTGAAGCAAAATATGCGGCGTACGAGATGATGATG 469
DB 310 CTGGGTACTCTGAGCTGCTTGAAGCAAAATATGCGGCGTACGAGATGATGATG 369
QY 470 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 529
DB 370 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 429
QY 530 GATATCAGCTCCCTCCACACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCA 589
DB 430 GATATCAGCTCCCTCCACACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCA 489
QY 590 GTGACACGACATTTGTATGAGAGGATGCTGATGACGACACTCAAGCTGTGCTC 649
DB 490 GCGCACGACATTTGTATGAGAGGATGCTGATGACGACACTCAAGCTGTGCTC 549
QY 650 CTGGCGCATGCGGCTCCATTCAGTTGTGAAATGCTCCACCTTAATCTTCTGCTC 709
DB 550 CAGTGCATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 609
QY 710 CTCACGACGAGTGGCTCCCTGCTAAAGGCGGCTCTCCGCTCATAGGCTGCTGCTG 769

DB 610 CCACAGAGCTGTGCCCCCTGGCTAACGGGGCTCCGCCCTGTATATCAACCTGCTG 669
QY 770 CGTTTGGCATCTGCGACCCACATGSCCAAGATTTCTTCTTACGAGATCTGTCCAG 829
DB 670 CATTTGCTCATCTGCGACCCACATGSCCAAGATTTCTTCTTATGAGATCTGTCCAG 729
QY 830 GGTACACATTTAAACATTAAGTAAAGGAGGACATCATTTGATGTGCGCCGCTCC 889
DB 730 GGTACACATTTAAACATTAAGTAAAGGAGGACATCATTTGATGTGCGCCGCTCC 789
QY 890 CAGACAGAAATGGGCTGTGCTGATCATCAAGCTGAATACAGGCGATTTATCA 949
DB 790 CAGTGCAGAGTGGCTGCTTCTGATCATCAAGTGAATATACAGGCGATTTATCA 849
QY 950 GCGACGACAAACATTAAGTGAACATTAAC-----AGTCCCCAGGCAAGAA 997
DB 850 GTCATGACAAACATTAAGTGAACATTAACAGGTTCTGTTAGTCCCAACCAAGAA 909
QY 998 CTATTTCTCATCATCAAGTTTACCCAGGCTGAGTGGCTTCAATATGAAATCTTTCTG 1057
DB 910 CTATTTCTCATCATCAAGTTTACCCAGGCTGAGTGGCTTCAATATGAAATCTTTCTG 969
QY 1058 ACATTGATCAAGATGAAACTCACTGACAGAAATTTATCTAGTATGACCTAATG 1117
DB 970 ACATTGATCAAGATGAAACTCACTGACAGAAATTTATCTGCAATGACCTCATG 1029
QY 1118 ATGTTCATATCTGTGTCAGCATGCGCCCGCTGCTCCAGATATCATCTCTCT 1177
DB 1030 ATGTTCATATCTGTGTCAGCATGCGCCCGCTGCTCCAGATATCATCTCTCT 1089
QY 1178 CTTTCAAGAGATTCGCTCCGCGAGTGGATGTCCTGATATGATCTTCTTCTGATG 1237
DB 1090 CTTTCAAGAGATTCGCTCCGCGAGTGGATGTCCTGATATGATCTTCTTCTGATG 1149
QY 1238 AGAGCTGCTGAGAGCTGCTGTCAGAGATGACAGCAGC---CAGAGAAAGTGC 1294
DB 1150 AGAGCTGCTGAGAGCTGCTGTCAGAGATGACAGCAGC---CAGAGAAAGTGC 1209
QY 1295 CTGTGACATTTGAAGATTAAGAGGGGAGAACTTCCAGCGAGCGAGTGTGAGAG 1354
DB 1210 CTGTGACATTTGAAGATTAAGAGGGGAGAACTTTCGAGCGAGCGAGTGTGAGAG 1269
QY 1355 AGCGCGCGAGCGCTCTTGGAGCAGCAGCAAGAGAGAGAGCGGTGGCTGAGCTG 1414
DB 1270 AGCGCGCGAGCGCTCTTGGAGCAGCAGCAAGAGAGAGAGCGGTGGCTGAGCTG 1329
QY 1415 AGCGCGCGAGCGAGAGAGAGAGAGAGCGGAGCGCGAGCAGAGCGCGAGC 1474
DB 1330 AGCGCGCGAGCGAGAGAGAGAGAGCGGAGCGCGAGCAGAGCGCGAGC 1389
QY 1475 TGGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1534
DB 1390 TGGAGCTGAG 1449
QY 1535 GAGAGAGAGAGATGAG 1594
DB 1450 GAGAGAGAGAGATGAG 1509
QY 1595 TTGAATGGAG 1654
DB 1510 TTGAATGGAG 1569
QY 1655 GCACCGTGTCTTGAAG 1714
DB 1570 ACATAGTTGTACTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1629
QY 1715 ACAAAAGAGATCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1774
DB 1630 ATAAAGAGATCAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1689
QY 1775 GCGAAGAAATTTGAG 1834
DB 1690 GCGAAGAAATTTGAG 1749

OY	1835	TACAGCAGAGTTGCAGGAATCTCAGCAAAATGCTTGAAAGACCTTATTCACAGAAACAGA	1894
Db	1750	TACAGCAACAATTACAGAAATCTCAGCAAAATCTTGAAAGACTTATTCAGAAAAACAGA	1809
OY	1895	TACTCAGTGACCGATTAAAAACAAGTCACAGCAAAACAGTTTGCAATGAGACTCGCTTCTTA	1954
Db	1810	TACTCAATGACCAATTTAAAAACAAGTTTCAGCAGAAACAGTTTGCACAGAGATTCACTTGTTA	1869
OY	1955	CCCTCAAAAAGCCCTTGGAAGCAAAAGAGAGCTGCCCCGACACAGTCCCGGAGCAGCTGG	2014
Db	1870	CACTTTAAAAAGCCTTTAGAGCAAAAAGAACTAGCTCGGACAGCACTTAGAACAACCACTGG	1929
OY	2015	ACGAGGTGAGAGAGAGACCAAGTCAAACTGCAGAGAAATGATGTTTCAACAACAGC	2074
Db	1930	ATGAAGTGAGAAAGAACTAGATCAAAAACCTACAGGAATGTTGATTTTCAATATATCAGC	1989
OY	2075	TGAAGGAACCTGAGAGAGATACCTAGCAAAACAGCACTTCACAGAGCAGAGTCCCTGAGG	2134
Db	1990	TGAAGGAACCTAGAGAAATATCACATATAGCAAACTCCAGAGCAAAAAGTCCATGAGAG	2049
OY	2135	CAGCGCACTGAAGAGAGAAAGAGCAGAGAGAGAGAGAGCTGAGATTAGAGAAAGCAAAAG	2194
Db	2050	CTGAAGCACTGAACAGAAACAGAAAGAACAGAACGAAAGATATAGAAATTAAGAAAACAAAAG	2109
OY	2195	AAGACGCTCAGAGACGAGTTCAAGAAAGGAGCAAGCAATGCGATGAGCATGTGCAGCAGG	2254
Db	2110	AAGAAAGCCCAAGAGCAGACTCAGGAAAGGAGCAAGCATGTGCTGGAGCATGTGCAGCAGG	2169
OY	2255	AG---GACACACACGCCCCCGGAAACCCCAAGAGAGAGACAGACTGAAGAGGGAACACA	2311
Db	2170	AGAGACGACATACAGAGACCAAGAAAACTTCACGAAAGAGAAAAACTGMAAAGGAGAGAGA	2229
OY	2312	GTGTCAAGAGAGAGAGGCGGAGAGAGAGAGCAAGCCGGAATATGCAAGACAAGCAGAGTC	2371
Db	2230	GTGTCAAAAAGAGAGATGGCGAGAAAAAAGGCAAAACAGGAAGCAACAGACAGCTGGTTC	2289
OY	2372	GGCTTTTTCATTCGCGATCAGAGAGCCAGCTTAAGCTGGCCAACCGAGCACCTTGCTCTACA	2431
Db	2290	GGCTTTTTCATCAACACCAAGAACCGAGCTTAAGCCAGCTGCAGGCAACCTCGTCTCACTG	2349
OY	2432	CAGAGAAAGGCCCGCTTACCAATTTGTGCACAGAGAGAGTGTAAAGTGGTATATTATCCGAG	2491
Db	2350	CAGAAAAAGGTTCACCTTACCAATTTGTGCACAGAAAAATGTAAAGTGGTATATTATCCGAG	2409
OY	2492	CGCTGTACCCCTTTGGAATCCAGAAAGTCACATGAGATCACCATCCAGCCAGACAGATATAG	2551
Db	2410	CAGTGTACCCCTTTGGAATCCAGAAAGCCATGATGAAATCATATCCAGCCAGAGACATAG	2469
OY	2552	TCAATGTGTGAATGAAGCCAGACTGTGAGAGCCAGATGAGCTTTGGAGAGAGCTGAAAGGGA	2611
Db	2470	TCATGTGTGAATGAAGCCCAAACTGTGAGAACCCGCTGTGGAGAGAAATTTAAAGGA	2529
OY	2612	AGACGGATGTGTTCCCTGTCAAACTATGACAGAAAAAGATTCCAGAAAAATGAGTTCCCACTC	2671
Db	2530	AGACAGGGTGTGTTCCCTGTCAAACTATGACAGAAAAATCCCAAGAAAAATGAGTTCCCGCTC	2589
OY	2672	CAGCCAAACCAGTGACCGATCTGACATCTGCGCCCTCGCCCAAACTGGCTCTGCGTAGA	2731
Db	2590	CAGTGAAACCAGTGACTGATTTCAACATCTGCGCCCTCGCCCAAACTGGCTCTGCGTAGA	2649
OY	2732	CCCGTGTCTCTTTGGCAGTGAACCTTTGTAGAGCCTCCACAAACCCCAAACTGGGCGAG	2791
Db	2650	CCCCCGCCCTTTGGCAGTGAACCTTTGTAGAGCCTCCACAGACCCCTTAATATACGTGGCGG	2709
OY	2792	ACTTCAGTTCACGTGAGCCACAGACTTCAAAACGAAAGCCAGAAACCGACCAACTCGGAGTA	2851
Db	2710	ACTTCAGTTCACGTGAGCCCAACAGCAAGATGAGAAACAGAAACCGACATTAATCGGAGTG	2769
OY	2852	CGTGGCGCGGTCAAGCTTCTCTGACCGTACCTAATGTCTGCGCAGTTACGAGAGATCAG	2911
Db	2770	CATGGGCAAGCCACGCTCTCTCAACGTTTCAAGTGGCGGCGCAGTTAAGGAGAGGTTCCG	2829

Qy	2912	CCCTTATACCCCAAGCAGCAGCTGAGCTCTCCCTCCATCTCCGATCCTTGAGCCAGGAGTGA	2369
Db	2830	CCTTATACCAAGCAGCAGCAGCTGAGCTCTCCCTCTCTCTGTGTAGGCAAGCTGA	2887
RESULT 11			
ID	AAK94139	standard; cDNA; 2131 BP.	
AC	AAK94139;		
DT	06-NOV-2001	(first entry)	
DE	Human full-length cDNA; SEQ ID NO: 2646.		
XX	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.		
XX	Homo sapiens.		
XX	EP1130094-A2.		
XX	05-SEP-2001.		
XX	07-JUL-2000; 2000EP-0114089.		
XX	08-JUL-1999; 99JP-0194486.		
XX	11-JAN-2000; 2000JP-0118774.		
XX	02-MAY-2000; 2000JP-0183765.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;		
XX	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;		
XX	WPI; 2001-524255/58.		
XX	P-PSDB; AAM93229.		
XX	830 Primers useful for synthesizing full length cDNA clones and their		
XX	use in genetic manipulation -		
XX	Claim 8; SEQ ID NO 2646; 1380bp + sequence listing; English.		
XX	The invention relates to primers for synthesizing full length cDNA		
XX	clones. 830 cDNA molecules encoding a human protein have been		
XX	isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA		
XX	molecules have been determined. Primers for synthesizing the full length		
XX	cDNA are useful for clarifying the function of the protein encoded by		
XX	the cDNA. The full length clones were obtained by construction of full		
XX	length enriched cDNA libraries that were synthesised by the oligo-capping		
XX	method. The primers enable the production of the full length cDNA easily		
XX	without any special methods. The present sequence is a full length		
XX	human cDNA of the invention.		
XX	Note: The sequence data for this patent did not form part of the printed		
XX	specification, but was obtained in CD-ROM format directly from EPO.		
XX	Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 other;		
Qy	Query Match	31.7%; Score 1610; DB 22; Length 2131;	
Qy	Best Local Similarity	85.5%; Pred. No. 0;	
Qy	Matches 1832; Conservative	0; Mismatches 295; Indels 15; Gaps 3;	
Db	1	GTGTCTCTCGTAGTACGGGGGCTCCGAGAAAGTATCCGAGCGGAGTCCGGAGCGA-- 57	
Qy	45	GTGTCTCTCGTAGTACGGGGGCTCCGAGAAAGTATCCGAGCGGAGTCCGGAGCGGAGTCCGGAGCGA-- 57	
Qy	105	GAGGCAAGCAGGCGGGGCGGGGATGTGTGCGCGGATCGGAGTCCGGAGCGGAGTCCGGAGCGA-- 57	
Db	58	-----CAGAGAGGGGGGGGAGATGTGTGCGGGGCTGCGGCTCTGTGCTCTCTCC 109	
Qy	165	-GCGGCGTGGGCGTGCACATGATTTGTGTGAGGGGGCGCGCGGACACCGGCGGAGAT 223	
Db	110	AGCGGCGCGTGGGCGGAGTATTTGTGCTGTGGGGCGGAGCGGAGCGGCGGAGAT 169	

QY 224 GAGGCGTCATCAGACGAGTGAACGTAATAGAACCATGGCTCAGTTTCCACACCTTTG 283
 DB 170 GAGGCGTCATCAGACGAGTGAACGTAATAGAACCATGGCTCAGTTTCCACACCTTTG 229
 QY 284 GTGGTAGCCTGAGATCTGGGCAATACGTGAGAGAAAGGGCAAGATGACACGAGT 343
 DB 230 GTGGTAGCCTGAGATCTGGGCAATACGTGAGAGAAAGGGCAAGATGATGACAGT 289
 QY 344 TCCCTTAGCCTGAGACCGGATAGGGGATTTATCTGCTGATACACGAGAACTTTT 403
 DB 290 TCCATAGTTTAAAGCCAAATCTGATTCATTCGTGATCMACTTGAACTTTT 349
 QY 404 TCCATCTGGGTTACCTCAGCTGTCTTAGCACAATATGGGCGCTAGCGGACATGATA 463
 DB 350 TTTCAATCTGGGTTACCTCAGCTGTCTTAGCACAATATGGGCACTAGCTGACGAA 409
 QY 464 ACGATGAAAGATGATCAAGTGAATTTTCCATAGCCATGAGCTTATCAACTGAAGC 523
 DB 410 ATGATGAAAGATGATCAAGTGAATTTTCCATAGCTATGAACTTATCAACTGAAGC 469
 QY 524 TACAAGATATCAGCTCCCTCCACACTTCCCTGTATGAAAGACGACCAAGTGTGTA 583
 DB 470 TACAAGATATCAGCTCCCTCCACACTTCCCTGTATGAAAGACGACCAATTTGTA 529
 QY 584 TTTCCAGTGCACACCATTTGGTATAGAGAGGATTGTAGCATGCGACCACTCAGCGTG 643
 DB 530 TTTCCAGTGCACACCATTTGGTATAGAGAGGATTGTAGCATGCGACCACTTACGCTG 589
 QY 644 TTTGCTCTGTGCGCAATGGGCTTCATTCAGTTGTGAATGTCTCACCTTATGATCTT 703
 DB 590 TTTGCTCTGTGCGCAATGGGCTTCATTCAGTTGTGAATGTCTCACCTTATGATCTT 649
 QY 704 CTGTCTCTCCAGCAGCAGTGGCTCCCTGGGCTTAAGGGGCTCTCCCTCATACAGCTC 763
 DB 650 CTGTCTCTCCAGCAGCAGTGGCTCCCTGGGCTTAAGGGGCTCTCCCTCATACAGCTC 709
 QY 764 TGCCCTGCTTGGCGATCCTGAGCAGCATGGGCAAGATCTTCTTACAGCAGATCTG 823
 DB 710 TGCCCTGATTTGCTATCTGAGCAGCATTTGCCAAAGATCTTCTTATGATGATCTG 769
 QY 824 GTCCAGGCTCAATTAATAACTAATAGTTACAGAAAGGCAATCATTCATGTGCGCAGC 883
 DB 770 GTCCAGGCTCAATTAATAACTAATAGTTACAGAAAGGCAATCATTCATGTGCGCAGC 829
 QY 884 CCCCCTCCAGCAGCAAAATGGGCTGTGCTCAGCTATCAAGGCTGAATAACGCGATAT 943
 DB 830 TCCCAACGAGTGGCAGATGGGCTGTCTCAGTCACTAAGACTGAATAACGCGCATAT 889
 QY 944 TCAACAGCCAGCAAAACTATGAGTGAACACTTAACAGGTCCCGCAGGCAAGACTATTC 1003
 DB 890 TCAATAGCTATGACAAACTATGAGTGAACACTTAACAGGTCCCGCAGGCAAGACTATTC 949
 QY 1004 TCATGCAATCAAGTTTACCCAGGCTCAGCTGCTTCAATATGAAATCTTTCTGACATG 1063
 DB 950 TTATGCACTCAAGTTTACAGAGGCTCAGCTGCTTCAATATGAAATCTTTCTGACATG 1009
 QY 1064 ATCAAGATGAGAAACTCACTGCAAGAAATTTATCTTACCTATGACCTAATGATGTG 1123
 DB 1010 ATCAAGATGAGAAACTCACTGCAAGAAATTTATCTTACCTAATGATGTGATG 1069
 QY 1124 CCATGTCTGTAGCAGCTGCGCGCGCTGCTGCTCCAGAAATACATCCCTCTTCTTCA 1183
 DB 1070 CTAATCTGTGCGCAACACTGCGCACTGTCTGCTCCGAAATACATTCACCTTCTTTTA 1129
 QY 1184 GAAAGATTCGCTCCGCGAGTGGATGTCCGTATACGCTCTTCTTCTGATGACAGGC 1243
 DB 1130 GAAAGATTCGATCTGCGAGTGGATATCTGTCACTAAGCTCAACATCTGATGACAGGC 1189
 QY 1244 TGCCGTAGGAGCGCTGCTCAGAGATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGA 1300
 DB 1190 TACAGAGAGAACGATTTTAAAGATGAAACAACAATTTGAAAGAAATTTACCTGTAA 1249
 QY 1301 CATTTGAAGATATAGAAAGCGGAGAACTTGAGCGAGGAGATGTGAGTGAAGAGCGCC 1360

DB 1250 CGTTTGAAGATATAGAGCGGAGAACTTTGACCTGTGCACTGGAACTGGAGAAACGAA 1309
 QY 1361 GCCAAGCGCTCTTGGAGCAGCAGCGCAAAAGCAGAGCGGTTGCTTCACTGAGCGCG 1420
 DB 1310 GGCAGGCTCTCTGGAACAGCAGCGCAAGAGCAGAGCGCTGCGCCAGCTGGAGCGGG 1369
 QY 1421 CCGAGCAGAGAGAAAGAGCGGAGCGCCGAGAGCAGAGAGCCAAAGCGGACGTGGAGC 1480
 DB 1370 CGAGCAGAGAGAGAGCGGAGCGCCGAGAGCAGAGAGCCAAAGAACCACTGGAGAC 1429
 QY 1481 TGGAGAAAGCAGCTGAGAGCAGCGGAGCTGAGCGGCAAGCAGAGAGAGAGAGAGA 1540
 DB 1430 TGGAGAGCACTGGAAGAAAGCAGCGGAGCTTAGAACGCGAGAGAGAGAGAGAGGGA 1489
 QY 1541 AGAGATCAGAGCGCGAGCGCCGAAACGGAACTGGAAGAGCAGCAGCACTTGAT 1600
 DB 1490 AAGAAATTTGAGAGCGCAGAGCGCTCAAAACGGGAACCTGGAAGGCAACGACACTTGAGT 1549
 QY 1601 GGGAAACGGAAACCGGAGCAGGAACTCCTGAATCAGAGAAACAGAGCAGAGGCGCACCG 1660
 DB 1550 GGGAAACGGAAATCGAAGCGAAGAACTTAATCAAGAAACAAAGAACAGAGACATAG 1609
 QY 1661 TGCTCTGGAAGCAGAGAGAGAACTCTGAGTTTGAATTGAGAGCTCTGAATGACAAA 1720
 DB 1610 TTGTACTGGAAGCAAAAGAAAGAACTTTGGAATTTGAAATTTGAAGCTCTTAATGATATA 1669
 QY 1721 AGCATCAGCTAGAGAGAAACTTCAAGATATCAGGTGTGCACTGCAACCCAGAGCGCAG 1780
 DB 1670 AGCATCAACTGAGAGAGAACTTCAAGATATCAGATGTGATGACCAACCCAGAGCGCAG 1729
 QY 1781 AAATTTGAGAGCAGAAAGCTAGAGAGCTAAGAAATTTGCTGAATACACCACTTACAGC 1840
 DB 1730 AAATTTGAGAGCAGAAAGCTAAGAGAGTTGAAGATTTGCCAATACCCATCTTACAGC 1789
 QY 1841 AGCAGTTGACAGAACTCTCAGCAAAATGCTTGAAGAACTTATCCAGAGAAACAGATACTCA 1900
 DB 1790 AACAAATTTACAGAACTCTCAGCAAAATGCTTGAAGAACTTATCCAGAGAAACAGATACTCA 1849
 QY 1901 GTGACAGTTTAAACAAGTCCAGAGAAACAGTTTGCATATAGAGACTCGCTTCTTACCTCA 1960
 DB 1850 ATGACCAATTTAAACAAGTTCAGAGAAACAGTTTGCACAGAGATTCATTTGTTACACTTA 1909
 QY 1961 AAAAGGCTTTGAGAGCAAAAGAGGTGCGCGGAGCAGCTCCGGAGCAAGCTGAGCAGG 2020
 DB 1910 AAAAGGCTTTGAGAGCAAAAGAACTAGCTGCGGAGCACTTACGAGCACTGATGATGAG 1969
 QY 2021 TGGAGAGAGAGACAGGCTCAAGCTGAGAGATTTGATTTTCAACAACAGCTGAGG 2080
 DB 1970 TGGAGAGAGAGACAGTATGATCAAAATCTTACAGAGATTTGATTTTCAATATACGCTGAGG 2029
 QY 2081 AACTGAGAGATATCATAGCAAAACAGCACTCCAGAAAGCAGAGGTCCCTGAGGCGCAGCGC 2140
 DB 2030 AACTGAGAGAAATCACAAATAGCAACAACCTCCAGAGCAAAAGTCCATGAGCGCTGAGC 2089
 QY 2141 GACTGAGCAGAAAGACAGAGAGAGAGAGCGCTGAGGATGAG 2182
 DB 2090 GACTGAGAAACAGAAAGAAACAGAAAGATCATGAAATGAG 2131

RESULT 12
 AA234573
 ID AA234573 standard; cDNA; 2079 BP.
 AA234573;
 AC
 AC
 AC
 DT 01-FEB-2000 (first entry)
 XX
 XX
 DE Human SH3D1A cDNA clone 5.
 XX
 XX SH3D1A gene; human; Down's syndrome; leukemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neutral disorder; thrombocytopenia;


```

QY 1381 CAGCGCAAAAGCAGAGCGGTTGGCTCAGCTGGAGCGCCCGCCAGCAGAGAGAAAG 1440
DB 1261 CAGCGCAAGAGCAGAGCGGCTGGCCAGCTGGAGCGGCGAGAGAGAGAGAG 1320
QY 1441 CCGGAGCGCCAGCAGAGCAGAGCGGCGAGCTGGAGCTGGAGAGCAGCTGGAGAG 1500
DB 1321 CGTAGCGCCAGAGCAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1501 CAGCGGAGCTGGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1381 CAGCGGAGCTAGAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1561 GCCCGCAAAACGGGAACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1441 GCTGCAAAACGGGAACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1621 GAACTCTGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1501 GAACTACTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1681 AAGACTGAGAGTTGAGTGAAGAGCTGTAAGAGAGAGAGAGAGAGAGAG 1740
DB 1561 AAGACTTGGAGTTGAGTGAAGAGCTTAATATGATAGAGAGAGAGAGAG 1620
QY 1741 CTTGAGAGTATCAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1621 CTTGAGAGTATCAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1801 TCTGAGAGAGTGAAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1681 TCTGAGAGTGAAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1861 CAATGCTTGGAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1741 CAATGCTTGGAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1921 CAGAGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1801 CAGAGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1981 GAGCTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1861 GAGCTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 2041 AAGCTGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 1921 AAGCTGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 2101 AAACGAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 1981 AAGCAAGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2161 GAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
DB 2041 GAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2078

```

RESULT 13
AAH16578
ID AAH16578 standard; cDNA; 2131 BP.

AAH16578;
26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:15658.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

```

XX 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
PF 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
PR 11-JUN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15658; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;

```

Query Match 27.6%; Score 1404.2; DB 22; Length 2131;

Best Local Similarity 81.0%; Pred. No. 0; Mismatches 283; Indels 126; Gaps 4;

Matches 1740; Conservative 0;

```

QY 7 GAGGAGAGTGGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66
DB 105 GAGGAGAGTGGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 164
QY 67 GCGAAGGAGAGATCCGAGAGCGGCTCCGAGAGCGGCGGAGAGAGAGAGAGAG 126
DB 165 GCGAAGGAGAGATCCGAGAGCGGCTCCGAGAGCGGCGGAGAGAGAGAGAGAG 213
QY 127 GGGATGGTGGCGGCGGCTGGAGATCGGCTTCTCCGCGGAGTGGCGGCTGACTGA 185
DB 214 GGGATGGTGGCGGCGGCTGGAGATCGGCTTCTCCGCGGAGTGGCGGCTGACTGA 273
QY 186 TTTGTGAGAGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 245
DB 274 TTTGTGAGAGGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 333
QY 246 ACGTAATAGAACATGAGCTAGTTCCACACTTTGAGTGGAGAGAGAGAGAGAGAG 305
DB 334 AAGTAACAGAACATGAGCTAGTTCCACACTTTGAGTGGAGAGAGAGAGAGAGAG 393

```

QY 306 CATACCTGTGAGAAAGGGCCAGCATGACGAGCTTCCTTAGCCTGAGCCGATAGC 365
 DB 394 CATACCTGTGAGAAAGGGCCAGCATGATGAGAGCTTCATAGTTTAAAGCAATATC 453
 QY 366 GGGATTATTATCTGTGATCAGCGAGAACTTTTTTTCATCTGGGTTACCTGACC 425
 DB 454 TGGATTATCTAGCGATCAAGCTAGAACTTTTTTTCATCTGGGTTACCTGACC 513
 QY 426 TGTCTTAGCAAAATATGGGCGCTAGCGGACATGAATTAAGAGAGATGATCAAGT 485
 DB 514 TGTCTTAGCAAAATATGGGCGCTAGCGGACATGAATTAAGAGAGATGATCAAGT 573
 QY 486 GGAATTTTCCATAGCCATGAGCTTATCAAACTGAAGCTACAGATATCAGCTCCCTC 545
 DB 574 GGAATTTTCCATAGCCATGAGCTTATCAAACTGAAGCTACAGATATCAGCTCCCTC 633
 QY 546 CACACTTCCCCCTGTATGAAACAGCAACCGAGTGTATTTCCAGTGCACAGCAATTTG 605
 DB 634 TGCACCTCCCTGTATGAAACAGCAACCGAGTGTATTTCCAGTGCACAGCAATTT-- 691
 QY 606 TATAGAGGATGTGATGATGACGACCACTCAGAGCTGTCTCTGTGCAATGGGCTC 665
 DB 692 ----- 691
 QY 666 CATTCAGTGTGGAATGTCTCACCCCTTAGTATCTTGTCCCTCCAGCAGAGCTGCC 725
 DB 692 -----GCACTGTGCC 702
 QY 726 TCCCTGTGCTAACGGGGCTCTCCGTCATACAGCCTGTGCTGCTTGTGCGATCTCG 785
 DB 703 CCCCCTGGCTAACGGGGCTCTCCGTCATACAGCCTGTGCTGCTTGTGCTATCTCTCG 762
 QY 786 AGCCATATGGCCAAAGATTTCTCTTCAGCAGATCTGTCTCCAGGCTCACAATTAAAC 845
 DB 763 AGCCATATGGCCAAAGATTTCTCTTCAGCAGATCTGTCTCCAGGCTCACAATTAAAC 822
 QY 846 TAAATTACAGAAAGGCAATCATTCGATGTGCGAGCGCCCTCCAGCAGAGATGGG 905
 DB 823 TAAATTACAGAAAGGCAATCATTCGATGTGCGAGCGCCCTCCAGCAGAGATGGG 882
 QY 906 TGTGCTCAGTCATCAAGGCTGAATTAACAGCAGTTATTAACAGCAGCAGCAAACTAT 965
 DB 883 TGTGCTCAGTCATCAAGGCTGAATTAACAGCAGTTATTAACAGCAGCAGCAAACTAT 942
 QY 966 GAGTGAACCTTAACAGGCTCCCGAGCAAGATTTATTCATGCAATCAAATTACCCA 1025
 DB 943 GAGTGAACCTTAACAGGCTCCCGAGCAAGATTTATTCATGCAAGTTACCA 1002
 QY 1026 GGCTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGAGAAATCTGCTG 1085
 DB 1003 GGCTCAGCTGGCTTCAATATGAAATCTTCTGACATTTGATCAAGATGAGAAATCTGCTG 1062
 QY 1086 AGAAGAAATTTATCTAGCTATGACCACTAAATGATGTGCAATGTCTGACCACTGCC 1145
 DB 1063 AGAAGAAATTTATCTAGCTATGACCACTAAATGATGTGCAATGTCTGACCACTGCC 1122
 QY 1146 GCGCGTCTGCTCCAGAAATCATCCCTCTCTTCAAGAGAGTTGCTCGGCACTGG 1205
 DB 1123 ACCGTGCTGCTCCAGAAATCATCCCTCTTCTTGAAGAGTTGATCTGGGAGTGG 1182
 QY 1206 GATGTCGCTCATTAAGCTCTTCTCTGATGATGAGAGGCTGCGAGAGCCGCTGCA 1265
 DB 1183 TATATCTGCTCATTAAGCTCTTCTGATGATGAGAGGCTGCGAGAGCCGCTTGA 1242
 QY 1266 GAGTGAACGAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGA 1322
 DB 1243 AGATGAACAACAACATTTAGAAAGAAATTAAGTTGATGAGATTAAGAGCGGA 1302
 QY 1323 GAACTTGAAGCGAGGAGGTGTGAGCTGAGAAAGCGCCGCAAGCGCTCTTGGAGCAGCA 1382
 DB 1303 GAACTTGAAGCGAGGAGGTGTGAGCTGAGAAAGCGAGCAAGCGCTCTTGGAGCAGCA 1362

QY 1383 GCGCAAGAGCAGAGACGGTGTGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAAAGCG 1442
 DB 1363 GCGCAAGAGCAGAGACGGTGTGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAGAGCG 1422
 QY 1443 GAGCGCCAGAGCAGAGAGCCAGCGGAGCTGTGAGAGAGAGAGAGAGAGAGAG 1502
 DB 1423 TGAAGCGCCAG 1482
 QY 1503 GCGGAGCTGAGAGCGCCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
 DB 1483 GCGGAGCTGAG 1542
 QY 1563 CCGCAAAACGGGAATCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1622
 DB 1543 TCGCAAAACGGGAATCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1602
 QY 1623 ACTCTGAATCAG 1682
 DB 1603 ACTCTGAATCAG 1662
 QY 1683 GACTCTGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1742
 DB 1663 GACTCTGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1722
 QY 1743 TCAGATATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1802
 DB 1723 TCAGATATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782
 QY 1803 TAGAGAGTGAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1862
 DB 1783 TAGAGAGTGAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1842
 QY 1863 AATGCTGGAAG 1922
 DB 1843 AATGCTGGAAG 1902
 QY 1923 GCAGAGAGTGTGAG 1982
 DB 1903 GCAGAGAGTGTGAG 1962
 QY 1983 GCTGCGCCGAG 2042
 DB 1963 ACTAGCTGCGAG 2022
 QY 2043 GCTGAG 2102
 DB 2023 ACTAG 2082
 QY 2103 ACAGAGAGTGTGAG 2151
 DB 2083 GCAGAGAGTGTGAG 2131

RESULT 14
 AA34574
 ID AA234574 strand: cDNA; 3231 BP.
 XX AA234574;
 AC
 AC
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA clone 9.
 XX
 KW SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopoenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KW
 OS Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers

FT CDS 2..1927
 FT /tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99MO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX Korenberg JR, Chen X;
 DR WPI; 1999-633829/54.
 DR P-PSDB; AA32158.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 PS Claim 2; Fig 14; 99pp; English.
 CC This is the nucleotide sequence of a non-full-length cDNA (clone
 CC 9) corresponding to a novel human SH3 gene, termed the SH3D1A
 CC gene, that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukaemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions, methods
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SQ Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;
 Query Match 27.4%; Score 1395.4; DB 20; Length 3231;
 Best Local Similarity 70.7%; Pred. No. 0;
 Matches 2359; Conservative 0; Mismatches 506; Indels 470; Gaps 19;

Db 241 AGACCACTGATGATGAGGAAAGAACTAGATCAAAAATCAAGAGATTGATATTTT 300
 Qy 2064 CAACAACCACTGAAAGAACTGAGAGATTCATAGCAAAACGACATCCGAAACGAG 2123
 Db 301 CAATATATCAGCTGAAGAACTTAAGAAATACAAATATAGCAACCTCCAGAGCAAAA 360
 Qy 2124 GTCCCTGAGGACCGGAGCTGAAGCGAAAGACAGAGAGAGAGAGCCGCTGAGTTAGA 2183
 Db 361 GTCCATGAGGCTGAACCACTGAACAGAAAGAAACAAAGACAAAGATCATGAAATTAGA 420
 Qy 2184 GAACAAAGAGAGACGCTCAGAGACGATTCAGAGAAAGGACAAAGATGCTGAGCA 2243
 Db 421 AAAACAAAAGAGAGAGCCAAAGACAGAGCTCAGAGAAAGGACAAAGAGCTGGTGAACA 480
 Qy 2244 TGTGACAGCAGAG---GAGCAGCCACGCCCCCGGAAACCCCAAGAGAGACAGACTGAA 2300
 Db 481 TGTGACAGCAGAGACGAGCAGATCAGAGACCAAGAAACTCCACAGAAAGGAAAACTGAA 540
 Qy 2301 GAGGAAAGACAGTGTCAAGAGAAAGAGAGGCGGAGAGAGAGCCAAAGCCGGAATGCAAGA 2360
 Db 541 AAGGAGAGAGAGTGTCAAAAAGAGAGTGGCAGAGAAAGGCAAAACAGAAACCAAGA 600
 Qy 2361 CAAGCAGAGTGGCTTTTCCATCCGATCAGAGACCAAGCTAAGCTGACCAACGAGCACC 2420
 Db 601 CAAGCTGGGTGGCTTTTCCATCAACCAAGAAACAGCTAAGCAGCTGTCCAGGACACC 660
 Qy 2421 CTGGTCTACACAGAGAAAGGCGGCTTACATTTCTGACAGAGAGTGTAAAAGTGT 2480
 Db 661 CTGGTCCACTCAGAAAAAGGTGTCACATTTCTGACAGAGAAATGTAAAAGTGT 720
 Qy 2481 ATATTACGAGCGGTGATACCCCTTGAATCCAGAAAGTCAGATGAGATCACATCCAGCC 2540
 Db 721 GTATTACGGGACCTGTACCTTTGAATCCAGAAAGCATATGAAATCATATTCACACC 780
 Qy 2541 AGAGATATAGTCAATGTGTATGTAAGCAAGCACTGAGAGCCAGAGATGGCTTTGAGAGGA 2600
 Db 781 AGAGACATATCATGTGTGATGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 840
 Qy 2601 GCTGAAAGGAAAGCGGATGTCTCTGCAAACTATGACAGAAAGATTCCAGAAATGA 2660
 Db 841 ATTAAGAGAAAGAGAGGTGTGTCTCTGCAAACTATGACAGAAATGCCAGAAATGA 900
 Qy 2661 GGTTCACATCCAGCAAAACAGTGAAGCGATCTGATCTGCTGCTGCTGCTGCTGCTGCTG 2720
 Db 901 GTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 Qy 2721 TCTGCTGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2780
 Db 961 CTGCTGAGAGACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
 Qy 2781 CAATGGGACAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2840
 Db 1021 TAACTGGGCGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
 Qy 2841 CAATGGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2900
 Db 1081 TAACTGGGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 Qy 2901 GCAGGATCAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2960
 Db 1141 GCAAGGCTGCTGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200
 Qy 2961 CGAGGTAAGAAAGGTGAGAGGCTTCAAGCGCAAGCCCTGTATCTCTGAGAGCCAAAAA 3020
 Db 1201 CGAGGTAAGAAAGGTGAGAGGCTTCAAGGCTCAAGCCCTATATCTCTGAGAGCCAAAAA 1260
 Qy 3021 AGACAAACACTTAATTTTAAAGAAAGTGAAGTATATACGCTTCTGAAACGCAAGACAT 3080
 Db 1261 AGACAAACACTTAATTTTAAAGAAAGTGAAGTATATACGCTTCTGAAACGCAAGACAT 1320
 Qy 3081 GTGTGCTTGAAGAGTTCAAGGTCAGAAAGGTTGTTCCCAAGCTTACGTAACAT 3140

Db 1321 GTGTGGTTTGGAAGTTCAAGTCAAGAGGGTGGTTCGCCAAGTCTTACGTGAACCT 1380
QY 3141 CATTTCAGGGCCCGTAAAGAAATCCACAGCATCGATCTGGCCCTACTGAAGTCTCTGC 3200
Db 1381 CATTTCAGGGCCCGTAAAGAAATCCACAGCATCGATCTGGTGTCTTCAAGAGTCTCTGC 1440
QY 3201 TAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGCCATTCCCGAGAAAGATTAT 3260
Db 1441 TAGCTTAAAGAGAGTGGCTTCTCCAGCAGCCAAAGCCGCTGTTCCGAGAA----- 1492
QY 3261 TGCATGTACACATACGAGAGTTCTGAGCAAGAGATTTAACCTTTCAGCAAGGGAGT 3320
Db 1493 ----- 1492
QY 3321 GATTGTGTTACCAAGAAATGTGTACTGTGTGACGGGACGGATGGGCGACAATGCCG 3380
Db 1493 ----- 1492
QY 3381 AGTCTTCCCTTCTAATATGTAGAGCTTAAAGATTCAAGGGCTCTGAACTGTGGAA 3440
Db 1493 ----- 1492
QY 3441 AACAGGAGTTTAGAAAAAAACCTGAATTTGCCCAAGTTATGCTTCTCACTGCTAC 3500
Db 1493 -----GAAATTTGCCAAGTTATTGCTCATACACCGCAC 1527
QY 3501 TGTGTCCGAACAATCACTGCTGCTCTGTGGGAGCTGTATCTGTATCCGAAAAAGAACCC 3560
Db 1528 CGGCCCCGAGACCTCACTGCGCCCTGTGAGCTGTATTTGATCCGAAAAAGAACCC 1587
QY 3561 AGGTGATGGTGGGAAAGAGAACTGCAGAGCTCGAGGAAAAAGCCAGATAGGTGTT 3620
Db 1588 AGGTGATGGTGGGAAAGAGAGCTGCAGAGCACTGGAAAAAGCCAGATAGGTGTT 1647
QY 3621 TCCAGCAATTTATGTCAAACTTAAAGCCCGGAAACAGCAAAATCAACCCCACTAGCT 3680
Db 1648 CCAGCTAATTTATGTAAAGCTTAAAGCCCTGGGAGAGCAAAATCACTCAAGAGCC 1707
QY 3681 ACCCAAGACCGGAGTGCAGCAGCAGTGTGCCAGGTGATGGGATGTACATTAACCGC 3740
Db 1708 ACCTAAGTCAACAGCATTAAGCCGAGTGTGCAGGTGATGTACATTAACCGC 1767
QY 3741 CCAGACAGTACGAACTAGCTTTCAGCAAAAGCCAGATCATCAAGTCTCTCAAGAG 3800
Db 1768 GCGAAATGACATGAGCTGCTTCAACAGGCGCAGATCATCAAGTCTCTCAAGAG 1827
QY 3801 GGAACCGGACTGTGTGAAAGAGAGTCAAGTGGGCAAGTGGCTCTTCCATCAATTA 3860
Db 1828 GGAACCGGACTGTGTGAAAGAGAGTCAATGACAAAGTGGGCTCTTCCATCAATTA 1887
QY 3861 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGATCATATGTTTCCATCCC 3920
Db 1888 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGATCATATGTTTCCATCCC 1947
QY 3921 CCTTCAGGCTTGAAGTCTCAAGAGAACCACTATCCATCATCTGAGCCAGAGGAGT 3980
Db 1948 CCTTCAGGCTTGAAG----- 1962
QY 3981 ATGGAGATGACACTTGATCATGTGACTTTCAGCATGATCACTTCTGAGTGA 4040
Db 1963 ----- 1962
QY 4041 GAAAGAACTACTGCAAGACAGTTTACCTCATTTGACCTTAGTTCATGTGAAATGT 4100
Db 1963 ----- 1962
QY 4101 CTGAGTCACTGCGTGACAGAGCAGAAATTCAGAACTGCAAGGATGATGAGTCC 4160
Db 1963 -----GTTCT 1967
QY 4161 TTTGGGCTTTCTTAGTCACTACAGTAC- CGGCCCCGCTTCAACAGGCGCTTTCAA 4219
Db 1968 TTTGGGCTTTCTTAGTCACTAAATTGACTTTCCCACTTTGACAGGTCCTTTCAA 2027

QY 4220 TAGTTTAAGATTATTTTAAATNGTATTTTAGCTTTTAAATTAATCTGAATGAT 4279
Db 2028 TAGTTTAAATTAATTTTAAATTAATTTTAGCTTTTAAATTAATTAATTAAT 2087
QY 4280 ACTTCTTGCTATTTTGGTTTAAACAAACACCCATATCAAGAGAGTCTGTGCGG 4339
Db 2088 GACTTCTTGCTATTTTGGTTTAAAGAAAGCCCATATCAAGAGATG-----CTGCAT 2142
QY 4340 ACGATTAAATGCTGTTCCGGGCGTACCGTAACTGAGAGCTTGTCTGATCTTTGCCGT 4399
Db 2143 GTGCTATTAATAATTTGTTCCAAATGTCATTAATCTGACATGTATGATTTTTCATTT 2202
QY 4400 TGTCCAGTGTCCCAACCAATTTGATTTGGGGCTGTCCCTG---CGGTAGACA 4455
Db 2203 TGTCCAGTGTACCAATTAATTTGACATTTGGGGCTTTTCCCTTACATAGAAATG 2262
QY 4456 CAGAGAGATGGGTGATCTGTTTGAATAATGTATGATGACTGAGCTGATAGAA 4515
Db 2263 CAGAGAGATGAGTATCTGTTTAAAGACGTATG- GAATGAGCCCAATTAAAGCAAG 2321
QY 4516 GGGGTTATGCTGTCTGTGACATCACTGTAC- TGTGCGCATGTACATCTGTACCG 4574
Db 2322 GTGTTGTGCTGTGTGTGTATGATGAGTGTACCTGTTGAGCATGTATTAATCTGTGA 2381
QY 4575 AAGAGTACT-CTTCTCTCATGCGTAAACCAACACCGTGTACATGCTCTCATACT 4633
Db 2382 CATTAAGAAATTAATGTTCTTCATGTGCAAGATTAATCTTGTGACATGCTCATATATAT 2441
QY 4634 --GCATTAATTTTACTTTCACAGATGACCTTGTAGGCACTGAGAAAGAC--CCATGTT 4690
Db 2442 TGCAATTAATTTTATTTTCAAGATGACCTTGTAGCACAATGAAAGACCTGTGTTT 2501
QY 4691 CCGTTTGTCTCAGATGATCCTAGTGTGCCCGTGTGTTTATTTTCAATCTGCG 4750
Db 2502 TGTTCGTCTCAGATTTATCTGTGTGATGTGTTTGTGTTTAAATTTTGC 2561
QY 4751 ATGTCTTACACCATTAATTAATGAGCCCACTGCCCCAGCGTTACATCATGAT 4810
Db 2562 GTGTTTGCATTAATTAATTAATGAGTACACAC- ACCATGAGGTGCTTACATCAACGAT 2620
QY 4811 CCACCGCTC-----TTAGTCTGTTAGCGTGAAG--TTTATCCAGTGTCTTTTATGAA 4864
Db 2621 TCCACGCTCTTTTATGCTCTGTACATGAAGTTTATTCAGTACTTTTCATGAA 2680
QY 4865 -----TATCTGAACAAGTATCTTGTGACAAAGAAAGATGATGAAGTCTCCCTGCA 4919
Db 2681 TGACCTATTTGACAAAGTATTTTCTTGACAAAGAAAGATGATGAAGTCTCCCTGCA 2740
QY 4920 ATTATTTTCCAGTGTTTTACATTTTAACTTGAATGATGTGGGGTGTCTACATTAAT 4979
Db 2741 ATTATTTT-CCAATGTTTACATTTTAACTTGAATGATGTGGGGTGTCTACATTAAT 2798
QY 4980 GAAATGGCTCTGCTGCGTGTGATGTTAACTTGTGCTGTAGCTGAAGCCGTGTG- 5037
Db 2799 GAAATGGCTCTGCTGCGTGTGATGTTT-AGATATGCTGTAGCTGAAGCCGTGTGT 2857
QY 5038 -CCTTAGATATTAAGTGAAGTGGGAGAGAAAT 5071
Db 2858 CTTTAAACATAGTTGGAAGCTCTCAATTAATAAT 2892

RESULT 15
AAS84762
ID AAS84762 standard; cDNA; 2874 BP.
XX
AC AAS84762;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #20566.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;


```
QY 2938 TCCTCCCATCTCCCGTCTGGGCGAGGTGAAAAGGTGAAAGGCTACAGCGCAAGCC 2997
    |||||
Db 1381 TCCTCCCGCTCTCTGTGCTAGGCGAGGTGAAAAGGTGAGGGGCTACAGAGCTCAAGCC 1440
    |||||
QY 2998 CTGTATCCCTGGAGAGCGAAAAAAGACACCACTTAATTTTAACAAAAGTGAAGTCATC 3057
    |||||
Db 1441 CTATATCTTGGAGAGCGAAAAAAGACCACTTAATTTTAACAAAATGATGTCATC 1500
    |||||
QY 3058 ACCGTTCTGGAAGAGACATGTGTGTTGGAAGTTCAGAGTCAAGGAGGTTGG 3117
    |||||
Db 1501 ACCGTTCTGGAAGAGACATGTGTGTTGGAAGTTC----- 1543
    |||||
QY 3118 TTCCCAAGTCTTACGTAAATCAATTCAGGCGCGTAAGAAATCCACAGATCATGAT 3177
    |||||
Db 1544 ----- 1543
    |||||
QY 3178 ACTGGCCCTAGTAGAAGTCTGTAGTAAAGAGATGGCTTCCCGCGCCAAAGCCA 3237
    |||||
Db 1544 ----- 1543
    |||||
QY 3238 GCCATTCGCGAGAGAGTTATTTGCAATGATACATAGAGAGTTCTGAGCAAGAGAT 3297
    |||||
Db 1544 -----AAGAAATTATGCGCATGTACACTTACGAGAGTTCTGAGCAAGAGAT 1590
    |||||
QY 3298 TTACCTTTCAGCAAGGAGATGATTTGTTACCAAGAAAGATGATGATGAGAG 3357
    |||||
Db 1591 TTACCTTTCAGCAAGGAGATGATTTGTTACCAAGAAAGATGATGATGAGAG 1650
    |||||
QY 3358 GGAACGATGGGCGACAGTCCGAGTCTTCTTCACTATGTGAGGCTTAAAGATTCA 3417
    |||||
Db 1651 GGAACAGTGGGCGACAGGCGGAGTCTTCTTCACTATGTGAGGCTTAAAGATTCA 1710
    |||||
QY 3418 GAGGGCTTGGAACTGTGGGAAAAACAGGAGTTTGAAGAAAACTGAAATTTGCCAG 3477
    |||||
Db 1711 GAGGGCTTGGAACTGTGGGAAAAACAGGAGTTTGAAGAAAACTGAAATTTGCCAG 1770
    |||||
QY 3478 GTTATGCTTCTTACGCTGCTACTGATCCGAACTCACCGTGGCTCTGGGAGCTG 3537
    |||||
Db 1771 GTTATGCTTCTTACGCTGCTACTGATCCGAACTCACCGTGGCTCTGGGAGCTG 1830
    |||||
QY 3538 ATTCTGATCGGAAAAAGAACCCAGGTGATGTGGGAAAGAGAACTGCAAGCTCGAGG 3597
    |||||
Db 1831 ATTTGATCGGAAAAAGAAC----- 1852
    |||||
QY 3598 AAAAAGCGCAGATAGGAGTGTTCAGCAATTAATGTCAACTCTAAGCCCGGAGACA 3657
    |||||
Db 1853 ----- 1852
    |||||
QY 3658 AGCAAAATCACCCCACTGAGTACCCAGAGCCGAGTGCAGCAGAGTGCAGG 3717
    |||||
Db 1853 -----CAAGTGCAGG 1866
    |||||
QY 3718 ATCGGATGTACGATTACACCGCCAGAACGATGACGAATGACTTACAGAAAGCCAG 3777
    |||||
Db 1867 ATCGGATGTACGATTACACCGCCAGAACGATGAGTGCCTTACAAAGGCGCAG 1926
    |||||
QY 3778 ATGATCAAGTCTCTCAAGAGAGAACCGGACTGTGAAAGAGAAAGTCAATGAGCAA 3837
    |||||
Db 1927 ATGATCAAGTCTCTCAAGAGAGAACCGGACTGTGAAAGAGAAAGTCAATGAGCAA 1986
    |||||
QY 3838 GTTGGGCTTTCATCCATCCAAATTAATTAAGCTGACCAAGATGAGACCCAGCAGCAA 3897
    |||||
Db 1987 GTTGGGCTTTCATCCATCCAAATTAATTAAGCTGACCAAGATGAGACCCAGCAGCAA 2046
    |||||
QY 3898 T 3898
    |||||
Db 2047 T 2047
```

THIS PAGE BLANK (USPTO)

Steven Jones, Jennifer Amano, Ian Bosgel, Aaron Butterfield, Susanna Chan, Readman Chu, Chris Fjell, Erin Garland, Ran Golin, Letticia Hsieh, Martin Krzyzwinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McInay, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacquelin Smith, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stottsch, Michael Thorne, Miranda Tsai, Matsuya van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Natasja Marris.

clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INTL. at: <http://image.llnl.gov>
Series: IRAL plate: 40 Row: n Column: 1
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796
This clone has the following problem: no cloning site / microdeletion.

FEATURES	Location/Qualifiers
source	1. .2126

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="caxon:9606"
/clone IMAGE:489011
/tissue_type="Pancreas, epithelioid carcinoma"
/clone_id="NH MGC 42"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"

```

Query Match	31.4%	Score 1598.2;	DB 11;	Length 2126;
Best Local Similarity	85.4%	Pred. No. 0;		
Matches 1803; Conservative	0;	Mismatches 304;	Indels 4;	Gaps 2;

Qy	92	CCGGAGACGGCCGGGAGGACGAGGAGCGGGCGGGAGATGCTGTGCGCGGCTCGGGACT	152
Db	14	CCGAGCGGGGCTCCGGGACGGACAGAGAGCGGGCGGGATGCTGTGCGGGCTCGGGCTC	73
Qy	152	CGGCGTTCCTCGC - GCGGGCTGCGGGCTTCACATGATTTGTGTGAGGGGCGGCCGCGCA	210
Db	74	CTGGGTCCCTCCAGGCGGGCGGTGAGCGGCACTGATTTGTCCCTGGGGCGCAGCGGGA	133
Qy	211	CCGGCCCGGAGATGAGGCGTCCATCAGCAAGTGAACTGAATAGAACCTAGGCTCACTTT	270
Db	134	CCGGCCCGGAGATGAGGCGGTCCATTTGCAAGGTAAAGAAACAGAACCATAGCTCACTTT	193
Qy	271	CCCAACCTTTGGGTGGTACCTCGATGTCGTGGGCAATPACGTGGAGGAAAGGGCCAG	330
Db	194	CCAAACCTTTTGTGTGACGCTGGATATCTTGGCCATPACTGTAGAGAAAGGCGAAG	253
Qy	331	CATGACGACAGTTCCTTAGCTGAAGCCGATAGCGGGATTTATTACTGTGATCAAGCG	390
Db	254	CATGATCAGAGTTTCCATAGTATTAAAGCCAAATPCTGGATTCTTACGTGGATCAAGCT	313
Qy	391	AGGAATTTTTTTTCCAACTGTGGGTTACTCGACCTGTCTTAGACAACAATATGGGCGCTA	450
Db	314	AGAAATTTTTTTTCAATCTGGGTTACTCAACTGTTTTTAGACAAATATGGGCACTA	373
Qy	451	GCGGACATGAATACGATGGAAGAGATGATCAAGTGGAAATTTTCCATGCGCATBAACCTT	510
Db	374	GCTGACATGAAATATGATGGAAGAAATGATCAAGTGGAGTTTTCCATAGCTATGAAACTT	433
Qy	511	ATCAAACTGAAGCTACAGGATATCAGCTCCCTCCACACTCCCTCGTATGAAACAG	570
Db	434	ATCAAACTGAAGCTACAGGATATCAGCTCCCTCTGCACTTCCCTCGTATGAAACAG	493
Qy	571	CAACCACTGGCTATTTCCAGTGCACCAAGCATTTGGTATAGAGGATTTCTAGCATGCCA	630
Db	494	CAACCACTGGCTATTTCTAGCGGACCAAGCATTTGGATAGGAGGATATGCCAGCATGCCA	553
Qy	631	CCACTCAGACTGTGCTCTGTGCCAAATGGGCTCATTTCAAGTTTGGAAATGTCTTCA	690
Db	554	CCGCTTACAGACTGTGCTCCAGTGGCCAAATGGGATCCATTTCAAGTTTGGAAATGTCTTCA	613

QY	691	CCCTTAGATCTTCGTCGCCCTCCAGACGAGATGCTCCCTCGGGCTTAACGGGGACTCTCC	750
Db	614	ACCTAGATCTTCGTCTGTTCCACAGCAAGCTGTGCCCTCGGGCTTAACGGGGACTCTCC	673
QY	751	GTCAATACAGCCTCTGCTCGCTTTGCGATCTCTCAGGCCACATGGCCAAAGATTCTTCC	810
QY	811	TTCAAGCAGATCTGGTCCAGGGTCAACATTAAACCTTAAGTTACGAAGGCAACAATCATTC	870
QY	734	TTTGTAGATCTGGTCCAGGGTCAACAATAAACCTAAATTACAAAGGCGCAGCTCATTT	793
QY	871	GATGTCGCGACGGCCCTCCAGACGAGAAAGGGCTGGCTCATGATCAAGGCTGAAA	930
Db	794	GATGTGCGCAATGTGCCACAGTGCGAGAGTGGCTGTTCTTCAGTCAATCAAGCTGAAA	853
QY	931	TACAGGCAAGTATTCAACAGCCAGACCAACAACTTAATGATGGACATTAAACGATCCCGAG	990
Db	854	TACAGGCAATTAATCAATAGTCATAGCAACAACTTAATGATGGACATTAAACGATCCCGAA	913
QY	991	GCAGAACTATTCTCATGCAATCAAGTTTACCCAGGGTCAAGCTGGCTTCAATATGAAAT	1056
Db	914	GCAAGAACTATTCTTATGCACTAGCTCAAGTTTACCAAGGGTCAAGCTGGCTTCAATATGAAAT	973
QY	1051	CTTTCTGACATTGATCAAGATGAAATCTCACAGAAAGATTTATCTGATGATGAC	1110
Db	974	CTTTCTGACATTGATCAAGATGAAATCTCACAGAGAAATTTATCTGCAATGAC	1033
QY	1111	CTAATGATGTTGTCATGTCGTCTGCTCAGCCATGCGCCCTGCTCCAGATATCATC	1170
Db	1034	CTCATTTATGTAAGCTATGTCTTGCCCAACACTGCGCACTGTCTGCTCCAGATATCATTT	1093
QY	1171	CCTCCTTCTTTCAGAAAGATTGCTCCGGCAGTGGAGTGTCCGTATTAAGCTTCTTCTT	1230
Db	1094	CCACCTTCTTTTGAAGAGATTCGATCTGCGCAGTGGTATATCTGTCAATAGCTCAACACTCT	1153
QY	1231	GTGATCATGAGAGCTGCTTGAGAGAACCGTCTCAAGATATGACAGCAGC---CAGAGAG	1287
Db	1154	GTAATCATGAGAGCTTACAGAGAAACCAAGTTTAAAGATGAACAAACAATTTGAAAAG	1213
QY	1288	AAACTGCTCTGACATTTGAAGATTAAGACCGGAGAACTTCGAGCGAGCGAGTGTGGAG	1347
Db	1214	AAATTTACTCTGTAAGCTTTGAAGATTAAGAGCGGAGAACTTTGAACGTGGCAACTGGAA	1273
QY	1348	CTGGAAGACGCGCGCCAAAGCGCTCTTGGAGCAGACCGCAAAAGACAGAGACGGTGGCT	1407
Db	1274	CTGGAGAAACGAAGGCAAGCTCTCTCGGAAACAGCAGCGCAAGAGACGCGCTGGCC	1333
QY	1408	CAGCTGGAAGCGCCCGAGCAGAGAGAGAAAGCGGGAGCGCCACAGACAGAGAGCCAGAG	1467
Db	1334	CAGCTGGAAGCGCGCGAGCAGAGAGAGAAAGAGCGGTGACCCACAGAGCAAGACCGCAA	1393
QY	1468	CGGCACTGGAAGCTGGAAGAACCACTGGAGAAAGCAGCGGAGCTGGAGCGGACAGAGAG	1527
Db	1394	AGACAACTGGAACCTGGAGAAACCACTGGAAAGAGCAGCGGAGCTTGAAGACGCAAGAGAG	1453
QY	1528	GAGGAGAGAGAAAGAGAGATCGAGAGGCGCGAGGCGCAAAACCGGAACTTGAAGAGCGAG	1587
Db	1454	GAGGAGAGAGAAAGAAATTGAAGAGGGAAGGCTGMAAACGGGAACCTTGAAGAGGAA	1513
QY	1588	CGACAACTTTGAATGGGAAACGGAACCGGAGACAGAACTCTGTAATCAAGAGAACAGAGAG	1647
Db	1514	CGACAACTTTGAGTGGGAAACGGAATCGAAGCGCAAAACCTAATTAATCAAAAGAAACAAAGAA	1573
QY	1648	CAGAGAGGCAACCGTGTCTCTGAAGGCAAGAGAAAGCTCTGGAGTTTGAATTTGAAGCT	1707
Db	1574	CAAGGAGCAATAGTTGTACTGAAGGCAAGAAAGAAAGATTGGAAATTTGAATTTGAAGCT	1633
QY	1708	CTGAATGCAAAAAGCATCAGCTGAAGAGAAACCTTCAGATATCAGGTGTGCACTGGCA	1767
Db	1634	CTTAATGATTAAGAAAGCATCACTGAGAGGAAACTTCAGATATCAGATGTGCAATTTGACC	1693
QY	1768	ACCCAGAGCAAGAAATTGAGAGCAAGAACAGTCTAGAGAGCTAAGATTGTCTGAATC	1827

```

Db      1694 ACCCAAGCGAAGAAATTTGAGACACAAACAAATCTAGAGATTGAGAAATGCGCAATC 1753
      1828 ACCCATTTACAGAGAGATTGACGAAATCTCAGCAAAATGTTGGAAGACTTATCCAGAG 1887
      1754 ACCCATCTACAGCAACATTTACGAGATCTCAGCAAAATGTTGGAAGACTTATCCAGAG 1813
      1888 AACAGATATCTCAGTACAGCAAGTTAAACAAATGTCAGCAGAAACAGTTGATAGAGACTG 1947
      1814 AACAGATATCTCAGTACAGCAAGTTAAACAAATGTCAGCAGAAACAGTTGATAGAGACTG 1873
      1948 CTTCTTACCTCAAAAGAGCTTTGAAAGCAAGAGCTGCGCCGAGAGACTCCGAGAG 2007
      1874 CTTGTATACCTTAAAGAGAGCTTTGAAAGCAAGAGACTGCGGAGAGACTTACAGAGAG 1933
      2008 CAGCTGAGAGAGGTGAGAGAGAGACAGAGTCAAACTGACAGAGATTGATGTTTCAAC 2067
      1934 CAACGTGAAGAGTGAAGAAAGAACTAGATCAAACTACAGAGATTGATGTTTCAAT 1993
      2068 AACCACTGAAGAGACTGAGAGAGATACATAGCAAAACAGCACTCCAGAGAGAGAGTCC 2127
      1994 AATCAGCTGAAGAGACTGAGAGAGATACATAGCAAAACAGCACTCCAGAGAGAGAGTCC 2053
      2128 CTGAGAGCAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2187
      2054 ATGAGAGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2113
      2188 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
      2114 AAAAAAAAAA 2124

```

RESULT 2

```

LOCUS      BC013578          2079 bp      mRNA      linear      HTC 04-MAR-2003
DEFINITION Homo sapiens, similar to interectin 1 (SH3 domain protein), clone
IMAGE:3878242, mRNA.
ACCESSION  BC013578
VERSION    BC013578.1 GI:15488896
KEYWORDS   HTC.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens

```

```

REFERENCE  1 (bases 1 to 2079)
AUTHORS    Strausberg, R.
TITLES     Direct Submission
JOURNAL    Submitted (04-SEP-2001) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA

```

```

REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
           Email: cgabbs-r@mail.nih.gov

```

```

           Tissue Procurement: DCTD/DP/Gazdar
           CDNA Library Preparation: Life Technologies, Inc.
           CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Sequencing Group at the Stanford Human Genome
           Center, Stanford University School of Medicine, Stanford, CA 94305
           Web site: http://www.shgc.stanford.edu
           Contact: (Dickson, Mark) mcd@paxil.stanford.edu
           Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
           R. M.

```

```

           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
           Series: IRAK Plate: 14 Row: C Column: 5
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 4504796
           This clone has the following problem: retained intron.

```

```

FEATURES
source     1..2079
           /organism="Homo sapiens"

```

```

BASE COUNT  703 a 451 c 492 g 433 t
ORIGIN
Query Match      31.1%; Score 1582; DB 11; Length 2079;
Best Local Similarity 85.8%; Pred. No. 0;
Matches 1780; Conservative 0; Mismatches 290; Indels 4; Gaps 2;

      129 GATGATGCGCGGCTGCGGAGTCCGCGCTTCTCTGCG-GGGGCTGCGGGCTGCACTGATT 187
      1 GATGATGCGCGGCTGCGGAGTCCGCGCTTCTCTGCGGCTTCCAGCGGCGCTGAGCGGACACTGATT 60
      188 TGTGTAGGGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 247
      61 TGTCCCTGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
      248 GTAATAGAACCATGCGCTCAGTTTCCACACCTTTGCGGTAGCCTGATGTCGCGCCA 307
      121 GTAACAGAACCATGCGCTCAGTTTCCACACCTTTGCGGTAGCCTGATGTCGCGCCA 180
      308 TAACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
      181 TAACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
      368 GATTATTACTGATGATCAAGGAGAGAACTTTTTCATCCATCTGGGTTACTCCGCGCTG 427
      241 GATTATTACTGATGATCAAGGAGAGAACTTTTTCATCCATCTGGGTTACTCCGCGCTG 300
      428 TCTTAGCACAATATAGGGGCTAGCGGAGATGATGATGATGATGATGATGATGATGATGATG 487
      301 TTTTAGCACAATATAGGGGCTAGCGGAGATGATGATGATGATGATGATGATGATGATGATG 360
      488 AATTTTCCATGACCATGAACTTATCAAACTGAACTGACAGATGATGATGATGATGATGATG 547
      361 AGTTTTCATGACCATGAACTTATCAAACTGAACTGACAGATGATGATGATGATGATGATG 420
      548 CACTTCCCCCTGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 607
      421 CACTTCCCCCTGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
      608 TAGAGAGAGATGCTAGATGCGCAGCAGCTACAGCTGTGCTCTGCTGCTGCTGCTGCTGCTG 667
      481 TGGAGAGATGCGCAGATGCGCAGCAGCTACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 540
      668 TTCCAGTTGTTGAATGTTCTCCACCCTTATGTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 727
      541 TTCCAGTTGTTGAATGTTCTCCACCCTTATGTTCTTCTGCTGCTGCTGCTGCTGCTGCTG 600
      728 CCTGGCTAAGGGGCTCTCTCCGTCATACAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 787
      601 CCTGGCTAAGGGGCTCTCTCCGTCATACAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
      788 CCACATGCGCAAGAGATTTCTCTTACAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 847
      661 CCACATGCGCAAGAGATTTCTCTTACAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
      848 AGTTACAGAGAGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 907
      721 AATTCAAAAGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
      908 TGCCTCAGATCAAGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 967
      781 TTCTCAGATCAAGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
      968 GTGACACTTAAACAGATCCCGAGGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTG 1027
      841 GTGACACTTAAACAGATCCCGAGGAGAGAGATGTTGATGTTGCTGCTGCTGCTGCTGCTG 900

```



```

Oy 1028 CTCAGCTGCTTCAATATGGAATCTTCTGACATTGATCAAGATGGAAGAACTCACTGAG 1087
Db 901 CTCAGCTGCTTCAATATGGAATCTTCTGACATTGATCAAGATGGAAGAACTTACAGAG 960
Oy 1088 AAGAAATTTATCTAGCTATGACCTTAATGATGTTGCCATGTCTGTCAGCCACTGCCG 1147
Db 961 AGGAATTTATCTGAGCAATGACCTCATTTGATGTATGTCTGAGCAACCACTGCCAC 1020
Oy 1148 CCGTCTGCTTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTGCTCCGAGTGGGA 1207
Db 1021 CTGTCTGCTTCCAGAAATACATCCCTCTTCTTGAAGAGTTGATCTGCACTGCTGTA 1080
Oy 1208 TGTCCGATATAGCTCTTCTTCTGGAATCAAGAGCTGCTGAGAGAGCCGTGTCAGAG 1267
Db 1081 TATCTGTATAGCTTCAACATCTGTATGATCAAGAGCTTCCAGAGAACCACTTTTAGAG 1140
Oy 1268 ATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAATTAAGAACCGGAGA 1324
Db 1141 ATGAAACAACAATTAAGAAAGAAATTAACCTGTAACCTTGAAGATTAAGAACCGGAGA 1200
Oy 1335 ACTTGAGCAGAGCAGTGTGAGCTGAGAGAACCCGCCAAGCCGCTTTGAGAGCAGC 1384
Db 1201 ACTTTGAACGTGGCAACCTGGAACCTGAGAGAAAGAAAGCAAGCTCTCTGGAACAGCAGC 1260
Oy 1385 GCAAGAGCAGAGCGGTTGGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAGAAAGCGG 1444
Db 1261 GCAAGAGCAGAGCGCTGCTGAGAGCGCCGAGCTGAGAGCGCCGAGCAGAGAGAGAGAGCTG 1320
Oy 1445 AGCCGAGAGCAGAGAGCGCAAGCGCAGCTGAGAGCTGAGAGAGCAGCTGAGAGAGCAGC 1504
Db 1321 AGCCGAGAGCAGAGAGCGCAAGCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCAGC 1380
Oy 1505 GGGAGCTGAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1564
Db 1381 GGGAGCTGAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGC 1440
Oy 1565 CAAAACGGGAACTGGAAGAGCGAGCAACTGGAATGGAAGAGAGAGAGAGAGAGAGAGAGC 1624
Db 1441 CAAAACGGGAACTGGAAGAGCGAGCAACTGGAATGGAAGAGAGAGAGAGAGAGAGAGAGC 1500
Oy 1625 TCTGGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1684
Db 1501 TACTGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Oy 1685 CTCTGAGTTGAGTTAGAGAGCTCTGATGACAAAAGAGATGAGTAAAGAGAGAGAGAGC 1744
Db 1561 CTCTGAGTTGAGTTAGAGAGCTCTGATGATGATGATGATGATGATGATGATGATGATG 1620
Oy 1745 AGGATATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1804
Db 1621 AGGATATCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Oy 1805 GAGAGCTTAAGATTTCTGAAATCACCACTTAACAGCAGAGAGTTCAGAGAGATTCAGCAA 1864
Db 1681 GAGAGCTTAAGATTTCTGAAATCACCACTTAACAGCAGAGAGATTCAGAGAGATTCAGCAA 1740
Oy 1865 TGCCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1924
Db 1741 TGCCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Oy 1925 AGAAGAGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1984
Db 1801 AGAAGAGTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Oy 1985 TGGCCCGCAGAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2044
Db 1861 TAGCTCGCAGAGAGCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Oy 2045 TGCAGAGAGATGATGTTTCAACAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2104
Db 1921 TACAGAGAGATGATGATGTTTCAACAATACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980

```

```

Oy 2105 AGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2164
Db 1981 AACCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Oy 2165 GGAAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
Db 2041 GAAAGATCATGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2074

```

RESULT 3

```

CA324450 818 bp mRNA linear EST 26-NOV-2002
LOCUS UI-M-FY0-ccl-a-05-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
DEFINITION IMAGE: 6820974 5', mRNA sequence.
ACCESSION CA324450
VERSION CA324450.1 GI:24542548
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

```

1 (bases 1 to 818)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL

```

COMMENT

```

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

```

```

The following repetitive elements were found in this cDNA
sequence: 338-455, >(GGA)n#Simple_repeat
Seq primer: PYX-5.
Location/Qualifiers

```

FEATURES

SOURCE

```

1. 818
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6820974"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FY0"
/note="Organ: Brain; Vector: PYX-Asc; Site 1:
Site 2: Not 1. The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into PYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCAGAGCAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Heimin Chin, Ph.D.,
program coordinator."

```

```

BASE COUNT 238 a 179 c 270 g 127 t 4 others
ORIGIN

```

```

Query Match 15 94; Score 809.8; DB 14; Length 818;

```

```

Best Local Similarity 99.34; Pred. No. 1.1e-166;
Matches 811; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Oy 1087 GAAGAAATTTATCTAGCTATGACCTTAATGATGTTGCGATGTCGTCAGCAGCACTGCCG 1146

```

```

Db      2  GAAGATTATCTTACTATGACCTAATGATGTCATGTCGTGTCAGCCAGCCG 61
      1147  CCGGTCTGCTGCTCAGAAATACATCCCTCTCTCTTCAAGAAAGTTGCTCCGCGACGTGG 1206
      62  CCGGTCTGCTGCTCAGAAATACATCCCTCTCTCTTCAAGAAAGTTGCTCCGCGACGTGG 121
      1207  ATGTCCGTATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAG 1266
      122  ATGTCCGTATTAAGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTCAGAG 181
      1267  GATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1326
      182  GATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 241
      1327  TTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1386
      242  TTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 301
      1387  AAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1446
      302  AAAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 361
      1447  CCGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1506
      362  CCGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 421
      1507  GAGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1566
      422  GAGCTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 481
      1567  AAAGCGGAACTGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAA 1626
      482  AAAGCGGAACTGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAAAGCGGAA 541
      1627  CTGAATCAGAGAAACAAGAGCAGAGAGCGGACCGGTGCTCTGAAGCAAGAGAAAGCT 1686
      542  CTGAATCAGAGAAACAAGAGCAGAGAGCGGACCGGTGCTCTGAAGCAAGAGAAAGCT 601
      1687  CTGAGATTGAGTTAAGCTCTGAATGACAAAAGATCAGCTAAGAGAAAGCTTCAAG 1746
      602  CTGAGATTGAGTTAAGCTCTGAATGACAAAAGATCAGCTAAGAGAAAGCTTCAAG 661
      1747  GATATCAGGCTGTCAGCTGGCAACCCAGAGCAAGAAATTGAGAGCAAGCAAGCTTGA 1806
      662  GATATCAGGCTGTCAGCTGGCAACCCAGAGCAAGAAATTGAGAGCAAGCAAGCTTGA 721
      1807  GAGCTAAGAAATGCTGAAATCAACCACTTAAGAGCAAGCTTGAAGCTTGAAGCTTGA 1866
      722  GAGCTAAGAAATGCTGAAATCAACCACTTAAGAGCAAGCTTGAAGCTTGAAGCTTGA 781
      1867  CTGGAAGACTTATTTCCAGAGAAACGATTAAGT 1903
      782  CTGGAAGACTTATTTCCAGAGAAACGATTAAGT 818

RESULT 4
LOCUS   BUS19029          906 bp      mRNA      linear      EST 12-SEP-2002
DEFINITION  AGENCOURT_10171971 NIH_MGC_134 Mus musculus cDNA clone
IMAGE:6516809 5', mRNA sequence.
ACCESSION  BUS19029
VERSION    BUS19029.1 GI:22826555
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.

```

```

FEATURES
source
  /organism="Mus musculus"
  /mol_type="mRNA"
  /db_xref="taxon:10090"
  /clone="IMAGE:6516809"
  /tissue_type="undifferentiated limb"
  /lab_host="DH10B (phage-resistant)"
  /clone_1ib="NIH MGC 134"
  /note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV, Site 2:
  NotI. Cloned unidirectionally. Primer: Oligo dT. Average
  insert size 1.7 Kb. Constructed by Resgen, Invitrogen
  Corp. Note: this is a NIH MGC library."
BASE COUNT      248 a      226 c      235 g      194 t      3 others
ORIGIN
Query Match      15.8%; Score 804.2; DB 13; Length 906;
Best Local Similarity 97.2%; Pred. No. 1.9e-165;
Matches 839; Conservative 0; Mismatches 20; Indels 4; Gaps 2;

      3354  GACGGGAACGCTGGGCGCAAGTCGAGCTTCCCTTAATATAGTAGAGCTTAAAG 3413
      1  GACGGGAACGCTGGGCGCAAGTCGAGCTTCCCTTAATATAGTAGAGCTTAAAG 60
      3414  TTCAAGGCGCTCTGAACTCTGGGAAAACAGGAGTTTGAAGAAAAAAGCTGAAATTG 3473
      61  TTCAAGGCGCTCTGAACTCTGGGAAAACAGGAGTTTGAAGAAAAAAGCTGAAATTG 120
      3474  CCAGTTATTGCTTCTTACGCTGCTACTGCTCCCAACAACTCACCTGCTCTTGCGCA 3533
      121  CCAGTTATTGCTTCTTACGCTGCTACTGCTCCCAACAACTCACCTGCTCTTGCGCA 180
      3534  GCTGATTCGATCCGGAAGAAAAGAACCCAGGTGATGTTGGAGAGGAACTGCAAGCTG 3593
      181  GCTGATTCGATCCGGAAGAAAAGAACCCAGGTGATGTTGGAGAGGAACTGCAAGCTG 240
      3594  AAGGAAAAAGCGCCAGATAGGGTGTTCACAGCAATTATGTCAACTTAAAGCCCGG 3653
      241  AAGGAAAAAGCGCCAGATAGGGTGTTCACAGCAATTATGTCAACTTAAAGCCCGG 300
      3654  AACCAAGCAAAATCAACCCCACTGAGCTTACCCAGACCGGACGTCAGCTGTCGA 3713
      301  AACCAAGCAAAATCAACCCCACTGAGCTTACCCAGACCGGACGTCAGCTGTCGA 360
      3714  GGTGATCGGATGTATGATTAACACCGCCAGAGCAATGAGCACTAAGCTTCAAGAAAG 3773
      361  GGTGATCGGATGTATGATTAACACCGCCAGAGCAATGAGCACTAAGCTTCAAGAAAG 420
      3774  CCAAGTATCAACGTCCTTCAACAGAGAGACCCCGATCTGTGAGAAAGGAAAGTCAAGT 3833
      421  CCAAGTATCAACGTCCTTCAACAGAGAGACCCCGATCTGTGAGAAAGGAAAGTCAAGT 480
      3834  GCAAGTTGGGCTCTTCCCATTTATTAAGCTGACCAACAGCATGAGACCCAGGCA 3893
      481  GCAAGTTGGGCTCTTCCCATTTATTAAGCTGACCAACAGCATGAGACCCAGGCA 540
      3894  GCAATGATCATATGTTGTCATCCCGCTCAGGCTTGAAGTCCCTCAAGAGAGACCCAC 3953
      541  GCAATGATCATATGTTGTCATCCCGCTCAGGCTTGAAGTCCCTCAAGAGAGACCCAC 600
      3954  TATCCCATATCACTGCCAGAGGAGTATGAGGAGTACAGCTTATATGATGACTTGA 4013
      601  TATCCCATATCACTGCCAGAGGAGTATGAGGAGTACAGCTTATATGATGACTTGA 660

```

QY 4014 GCATGATCATCTACGCTTCTGAGTAGAAGAACTCAGTCAGAGACAGTTTACCTCATTT 4073
 DB 661 GCATGATCATCTACGCTTCTGAGTAGAAGAACTCAGTCAGAGACAGTTTACCTCATTT 720
 QY 4074 GACCTTAGTGTGATGATCGAATATGTCGATCAGTCAGTCAGAGAGCAGAGCAATTT 4133
 DB 721 GACCTTAGTGTGATGATCGAATATGTCGATCAGTCAGTCAGAGAGCAGAGCAATTT 780
 QY 4134 GCAGACATGACACAGGCT--GGTGGGTCTTTTGGGG--CTTTCTAGTCACTCAGACTGA 4189
 DB 781 GCAGACATGACACAGGAGTGTGGGGGCTCTTTTGGGGGCTTTCATGACTCAAAACTG 840
 QY 4190 CCGGCGCGGCTTTCACAGCGGCG 4212
 DB 841 ACGGCGCGGCTTTCACAGCGG 863

RESULT 5
 B0942708 911 bp mRNA linear EST 21-AUG-2002
 LOCUS B0942708 5', mRNA sequence.
 DEFINITION B0942708.1 GI:22358186
 ACCESSION B0942708.1 GI:22358186
 VERSION B0942708.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 911)
 NIH-MGC http://mgi.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: Reggen, Invitrogen Corp.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/MLN at:
 http://image.llnl.gov
 Plate: LLM13743 row: n column: 03
 High quality sequence start: 10
 High quality sequence stop: 564.
 Location/Qualifiers
 1..911
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6314690"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 129"
 /note="Organ: olfactory epithelium; Vector: pCMV-SPORT6.1; cdb: Site 1: EcoRV; site 2: NotI. Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Reggen, Invitrogen Corp. Note: This is a NIH MGC library."
 BASE COUNT 271 a 210 c 280 g 150 t
 ORIGIN

Query Match 15.6%; Score 795.6; DB 13; Length 911;
 Best Local Similarity 98.0%; Pred. No. 1.5e-163;
 Matches 890; Conservative 0; Mismatches 9; Indels 9; Gaps 8;

QY 936 GCAGTTATTCACACGACCAACAACATATGATGAGACATTAACAGGTCGCCAGGCAAG 995
 DB 1 GCAGTTATTCACACGACCAACAACATATGATGAGACATTAACAGGTCGCCAGGCAAG 60
 QY 996 AACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCATATGGAATCTTTC 1055
 DB 61 AACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCATATGGAATCTTTC 120

QY 1056 TGACATTGATCAAGATGAAAAATCTACGACAGAGAAATTATCTAGCTATGACCTAAT 1115
 DB 121 TGACATTGATCAAGATGAAAAATCTACGACAGAGAA- TTAATCTAGCTATGACCTAAT 179
 QY 1116 TGATGTTGCCATGTCTGTGTCAGCCACTGCCGCCCTCTGCTCCAGAAATACCTCTCC 1175
 DB 180 TGATGTTGCCATGTCTGTGTCAGCCACTGCCGCCCTCTGCTCCAGAAATACCTCTCC 239
 QY 1176 TTCTTTGAGAAAGATTGGCTTCGGCAATGGGATGTCCTCATTAAGCTTTCTTCTGGA 1235
 DB 240 TTCTTTGAGAAAGATTGGCTTCGGCAATGGGATGTCCTCATTAAGCTTTCTTCTGGA 299
 QY 1236 TCAGAGGCTGCTGAGAGAGCCGTGTCAGAGGATGAGAGAGCAGAGCAAGAAACTGCC 1295
 DB 300 TCAGAGGCTGCTGAGAGAGCCGTGTCAGAGGATGAGAGAGCAGAGCAAGAAACTGCC 359
 QY 1296 TGTGACATTTGAGATGAAGAGCCGAGAACTTTCAGCAGAGCAGTGTGAGCTGAGAA 1355
 DB 360 TGTGACATTTGAGATGAAGAGCCGAGAACTTTCAGCAGAGCAGTGTGAGCTGAGAA 419
 QY 1356 GCGCGCCAGAGCGCTCTTTGAGAGCAGCCGAAAGACAGAGAGCGTTTGGCTCAGCTGA 1415
 DB 420 GCGCGCCAGAGCGCTCTTTGAGAGCAGCCGAAAGACAGAGAGCGTTTGGCTCAGCTGA 479
 QY 1416 GCGCGCCAGAGAGAGAGAGAGAGCGGAGAGCCGACAGAGAGAGGCAAGCGGAGCT 1475
 DB 480 GCGCGCCAGAGAGAGAGAGAGAGCGGAGAGCCGACAGAGAGAGGCAAGCGGAGCT 539
 QY 1476 GGAAGCTGAGAGAGAGAGCTGAGAGAGCAGCGAGCTGAGAGCGAGAGAGAGAGAG 1535
 DB 540 GGAAGCTGAGAGAGAGAGCTGAGAGAGCAGCGAGAGCTGAGAGCGAGAGAGAGAG 599
 QY 1536 GAGAGAGAGATCGAGAGAGCGCGAGAGCGGCAAAACGGAACTGGAAGAGCAGCAACT 1595
 DB 600 GAGAGAGAGATCGAGAGAGCGCGAGAGCGGCAAAACGGAACTGGAAGAGCAGCAACT 659
 QY 1596 TGAATGGGAAC-GGACCGGAGAGAGAGAGCTCTGAATCAGAGAGCAAGAGAGAGAG 1654
 DB 660 TGAATGGGAACCGGAGAGAGAGAGAGCTCTGAATCAGAGAGCAAGAGAGAGAG 719
 QY 1655 GC-ACCGTGTCTCTGAGAGAGAGAGAGAGCTCTGAGTTTGAAGAGCTCTGAA 1712
 DB 720 GCCACCGTGTCTCTGAGAGAGAGAGAGAGCTCTGAGTTTGAAGAGCTCTGAA 779
 QY 1713 TGAAGAAAAGATCAGCTAGAGAGAAACTCAGAGATTCAGGT-GTGAGTGGCAACC 1771
 DB 780 TGAAGAAAAGATCAGCTAGAGAGAAACTCAGAGATTCAGGTGTGAGTGGCAACC 839
 QY 1772 AGA-GGCAAGAAATTGAGAGAGAG-ACAAGTCTAGA-GAGCTAAGAAATTGC-TGAATC 1827
 DB 840 AGAAGGCAAGAAATTGAGAGAGAGCAAGAACTTAAAGAGCTTAAATTCCTTGAATC 899
 QY 1828 ACCCACTT 1835
 DB 900 ACCCACTT 907

RESULT 6
 AK082606 2589 bp mRNA linear HTC 05-DEC-2002
 LOCUS AK082606
 DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length
 enriched library, clone: C300688.04 product: Intersectin (SH3 domain
 protein 1A), full insert sequence.
 ACCESSION AK082606
 VERSION AK082606.1 GI:26349812
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, K., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamahata, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleschbaum, M., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarini, R., Barsh, G., Blake, J., Boileau, D., Bojunga, N., Carinacci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Flehner, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzei, J., Mombauts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Welter, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 12108560
PUBMED 12108560
REFERENCE 6 (bases 1 to 2589)
Adachi, J., Aizawa, K., Akimura, T., Aizawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnate, N., Okazaki, Y., Saito, R., Saitoh, H., Sekai, C., Sekai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission.
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-cho, Yokohama,

COMMENT Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://phantom.gsc.riken.go.jp/
Location/Qualifiers
1. 2589
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:C230068L04"
/db_xref="taxon:10090"
/clone="C230068L04"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1. 2269
/note="unlabeled protein product; interseotin (SH3 domain protein 1A) (MGD|MG1:1338069, GI|NM_010587, evidence: BLASTN, 99%, match=2267)
putative"
/codon_start=2
/protein_id="BAC38546.1"
/db_xref="GI:26349813"
/translation="PFRGPRVKSNSIDGTPESPASLKRVAAPAKPAIPGEFTAM
LYVESBGDLTFPOGQDVVYVVKDMDWGTGSPKSGVSPNRYVRLKDSGTGK
TSLKRPKRIANVAVATGTEQTLTAPQILIRKPNPGMBSGLARCKRQIG
INFPANVYKLSFGSKITPTELPTKAVQAPQVGMVDTAQNDELFSKQIIV
LNKPDQWPKSGVQVGLFSPNRYVLTIDMDPQQSCSDHLLMPTPKRQGY
HLIIVTEYVVDLQVTEIFQKPTESLITREKVAMFVAMKELIMCNILKLR
VKKSGEMPVKMGIDILASQIPNOPYRSCQNLGALIQKDEADPFEVK
RLAMPKMGKLSFILLKPMQRYVRLIKNITLLENPMENHDSHKLALKEVEL
CSQNVGKREKNSRLRIQAHVQCELSQVLYVNSVNTCLGPKFLSGKLYRKS
NKEIYGFNDPLLTQITKPGSSGTQVSPKSNLYKMTPTFLNEVLVLPD
PSSGDFPIHSHIDVYVLRASINERTAWQKIAASLYETEKKEKAYLVASQ
RATGIRGLRVNVEGIELKPSKHSNPKYCVWGSQCHITKTIQDTLNPKNVSC
FPIRLDEVLCTIVFERSDQSLADFLKTEIRVADIKKDGSKGVPVTKLLHVEPT
GEIVRYDLQLDFEB"
BASE COUNT 693 a 662 c 688 g 546 t
ORIGIN
Query Match 14.7%; Score 748.8; DB 11; Length 2589;
Best Local Similarity 99.7%; Pred. No. 3.5e-153;
Matches 750; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
3148 GGGCCCGTAAGAAATCCAGACATCGATCTGCGCTTCTAGTCTA 3207
14 GGGCCCGTAAGAAATCCAGACATCGATCTGCGCTTCTAGTCTA 73
3208 AAGAGAGTGGCTTCCCGCCGCGCCAGCCAGCCATTCGGAGAAAGTTATGGCAG 3267
74 AAGAGAGTGGCTTCCCGCCGCGCCAGCCAGCCATTCGGAGAAAGTTATGGCAG 133
3268 TACACATACGAGAGTTCTGAGACAGAGATTTAACTTTCACAGAGGGATGATG 3327
134 TACACATACGAGAGTTCTGAGACAGAGATTTAACTTTCACAGAGGGATGATG 193
3328 GTTACCAAGAAAGATGTGATCTGTGAGCGGAGACGTGGGGGACAGTCCGAGTCTTC 3387
194 GTTACCAAGAAAGATGTGATCTGTGAGCGGAGACGTGGGGGACAGTCCGAGTCTTC 253
3388 CCTTCTACATAGTGAAGTTAAATTCAGAGGCTCTGGAACCTGCTGGGAAACAGGG 3447
254 CCTTCTACATAGTGAAGTTAAATTCAGAGGCTCTGGAACCTGCTGGGAAACAGGG 313
3448 AGTTAGGAAAGAACTGGAATTCGCCAGTTATGCTTCTACGCTGCTAGTGGTCC 3507

Db 314 AGTTTGAAGAAAAAAGCTGAATTTGCGCAGGTTATTGCTTCTACGCTTACTGTTGCC 373

Qy 3508 GAACAACCTACCCCTGCTGCTGCGCAGCTGATTTGATTCGGAAAAAGAACCCAGGTGGA 3567

Db 374 GAACAACCTACCCCTGCTGCTGCGCAGCTGATTTGATTCGGAAAAAGAACCCAGGTGGA 433

Qy 3568 TGGTGGGAAAGGAACTGCAAGCTGAGGAAAAAGCCCGAGATGGGTGGTTTCACGA 3627

Db 434 TGGTGGGAAAGGAACTGCAAGCTGAGGAAAAAGCCCGAGATGGGTGGTTTCACGA 493

Qy 3628 AATTATGTCAACTTCTAAGCCCGGAAACAAGCAAAATCACCCCACTGAGCTAACCCAG 3687

Db 494 AATTATGTCAACTTCTAAGCCCGGAAACAAGCAAAATCACCCCACTGAGCTAACCCAG 553

Qy 3688 ACCGAGTGCAGCCGACGAGTGTGCGAGGTGATGCGGATGTACGATTACACCGCCAGAAC 3747

Db 554 ACCGAGTGCAGCCGACGAGTGTGCGAGGTGATGCGGATGTACGATTACACCGCCAGAAC 613

Qy 3748 GATGACGAACTAGCTTACGAAAGGCGCAGATCATCAAGTCTCAGACAGAGGAGCCG 3807

Db 614 GATGACGAACTAGCTTACGAAAGGCGCAGATCATCAAGTCTCAGACAGAGGAGCCG 673

Qy 3808 GACTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCATCCATTTATGTAAG 3867

Db 674 GACTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTCTTCCATCCATTTATGTAAG 733

Qy 3868 CTGACCAACAGATGAGACCCCGACGACCAATG 3899

Db 734 CTGACCAACAGATGAGACCCCGACGACCAATG 765

RESULT 7
CA750495 732 bp mRNA linear EST 27-NOV-2002
LOCUS UI-M-F70-cde-p-10-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
DEFINITION IMAGE: 6832091 5', mRNA sequence.
CA750495
ACCESSION CA750495.1 GI:25575732
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 732)
NIH-MGC <http://mgi.mci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 637-722, >(CAG)nSimple_repeat
Seq primer: pyx-5.

FEATURES

source
location/Qualifiers
1..732
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6832091"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1ib="NIH BMAP FY0"
/note="Organ: Brain; Vector: pyx- Asc; Site_1: Ecor I;

Site 2: Not I; The library was constructed according
Bonaldo, Lemon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to RNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGCAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): "Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."

BASE COUNT 197 a 187 c 209 g 139 t
ORIGIN
Query Match 14.3%; Score 727.2; DB 14; Length 732;
Best Local Similarity 99.6%; Pred. No. 1,4e-148;
Matches 729; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 788 CCACATGGCCAAAGAGTTCTTCTTCCAGCAGATGTGTTCCAGGTCACATTAACCTA 847
Db 1 CCACATGGCCAAAGAGTTCTTCTTCCAGCAGATGTGTTCCAGGTCACATTAACCTA 60
Qy 848 AGTTACGAAGGCGCAATCAATTGATGTCGCGAGCGCCCTCCAGACAGAAATGGGCTG 907
Db 61 AGTTACGAAGGCGCAATCAATTGATGTCGCGAGCGCCCTCCAGACAGAAATGGGCTG 120
Qy 908 TGCCCTCAGTCATCAAGGCTGAAATATACAGGAGTTATTCACAGCAGACAAACTATGA 967
Db 121 TGCCCTCAGTCATCAAGGCTGAAATATACAGGAGTTATTCACAGCAGACAAACTATGA 180
Qy 968 GTGACACTTAACAGGTCGCCAGGCAAGAACTATTCATATGCAATCAAGTTTACCACAG 1027
Db 181 GTGACACTTAACAGGTCGCCAGGCAAGAACTATTCATATGCAATCAAGTTTACCACAG 240
Qy 1028 CTCAGCTGGCTTCAATATGGAATCTTCTGATTCATTAAGATGAGAAATCTCACTGACG 1087
Db 241 CTCAGCTGGCTTCAATATGGAATCTTCTGATTCATTAAGATGAGAAATCTCACTGACG 300
Qy 1088 AAGATTTATCTTACTATGACCTTAATGATGTGTCATGTCTGTACGCACTGCGGC 1147
Db 301 AAGATTTATCTTACTATGACCTTAATGATGTGTCATGTCTGTACGCACTGCGGC 360
Qy 1148 CCGTCTGCTTCCGAATATATCCCTCTTCTTCAAGAGTTGCTTCGCGCAGTGGGA 1207
Db 361 CCGTCTGCTTCCGAATATATCCCTCTTCTTCAAGAGTTGCTTCGCGCAGTGGGA 420
Qy 1208 TGTCCGTCATAGCTCTTCTTCTGATGTCAGAGGCTGCTGAGAGCGCTGTCAGAGG 1267
Db 421 TGTCCGTCATAGCTCTTCTTCTGATGTCAGAGGCTGCTGAGAGCGCTGTCAGAGG 480
Qy 1268 ATGAGCAGCAGCAGAGAAAGAACTGCTGTGACATTTGAAGATGAAGCGGGAACCT 1327
Db 481 ATGAGCAGCAGCAGAGAAAGAACTGCTGTGACATTTGAAGATGAAGCGGGAACCT 540
Qy 1328 TCGAGCGAGGAGTGTGAGCTGAGAGAGCGCCCGCAAGCGCTTGTGAGAGCAGCGCA 1387
Db 541 TCGAGCGAGGAGTGTGAGCTGAGAGAGCGCCCGCAAGCGCTTGTGAGAGCAGCGCA 600
Qy 1388 AAGAGCAGGAGCGGTTGGCTCAGCTGAGAGCGCGCGAGAGAGAGAAAGCGGAGC 1447
Db 601 AAGAGCAGGAGCGGTTGGCTCAGCTGAGAGCGCGCGAGAGAGAGAAAGCGGAGC 660
Qy 1448 GCGAGAGCAGAGAGCCAAAGCGGAGCTGAGAGCTGAGAGACAGCTGAGAGAGCGGG 1507
Db 661 GCGAGAGCAGAGAGCGCAAGCGGAGCTGAGAGCTGAGAGACAGCTGAGAGAGCGGG 720
Qy 1508 AGCTGAGCGGC 1519
Db 721 AGCTGAGCGGC 732

RESULT 8
CB248849
LOCUS
DEFINITION
CB248849
IMAGE: 5719103 5', mRNA sequence.
ACCESSION
CB248849.1 GI:28387663
VERSION
EST.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 725)
NIH-MGC <http://mgi.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pyx-5.
Location/Qualifiers
1. 725
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 5719103"
/issue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1ib="NIH_BMAP_EXO"
/note="Organ: Brain; Vector: pyx-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is GTGGGTGGA. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP). 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 250 a 151 c 221 g 102 t 1 others
ORIGIN

Query Match 14.2%; Score 724; DB 14; Length 725;
Best Local Similarity 99.9%; Pred. No. 7.3e-148;
Matches 724; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1594 CTGTGATGGGAACGGAACCGGAGACGGAATCTCTGAATCAGAGGAACAGAGCGAGG 1653
DB 1 CTGTGATGGGAACGGAACCGGAGACGGAATCTCTGAATCAGAGGAACAGAGCGAGG 60
QY 1654 GGCACCGTGTCTCTGAAGCAAGAGAAAGACTCTGAGTTGAGTTAGAAAGCTCGAAT 1713
DB 61 GGCACCGTGTCTCTGAAGCAAGAGAAAGACTCTGAGTTGAGTTAGAAAGCTCGAAT 120
QY 1714 GACAAAGACATCAGTTAGAGAAATTCAGAGATATCAGGTGTGACGTGGCAACCCAG 1773
DB 121 GACAAAGACATCAGTTAGAGAAATTCAGAGATATCAGGTGTGACGTGGCAACCCAG 180

QY 1774 AGCAAGAAATTTAGAGACGGAACAGTCTAGAGAGCTAAGAATTGCTGAATCAACCCAC 1833
DB 181 AGCAAGAAATTTAGAGACGGAACAGTCTAGAGAGCTAAGAATTGCTGAATCAACCCAC 240
QY 1834 TTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGGAAAGACTTAATTCAGAGAAACAG 1893
DB 241 TTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGGAAAGACTTAATTCAGAGAAACAG 300
QY 1894 ATACTCAGTGCACCTGTAACAAAGTCCAGCAAAACAGTTTGCATAGAGACTTCCTTCTT 1953
DB 301 ATACTCAGTGCACCTGTAACAAAGTCCAGCAAAACAGTTTGCATAGAGACTTCCTTCTT 360
QY 1954 ACCCTCAAAGACCTTGGAAACAAAGAGAGTGGCCCGGACAGCTCCGGAGACAGCTG 2013
DB 361 ACCCTCAAAGACCTTGGAAACAAAGAGAGTGGCCCGGACAGCTCCGGAGACAGCTG 420
QY 2014 GACGAGTGGAGAGAGACGACAGTCAAAAGCTGACAGAGATTGATTTTCAACAAACAG 2073
DB 421 GACGAGTGGAGAGAGACGACAGTCAAAAGCTGACAGAGATTGATTTTCAACAAACAG 480
QY 2074 CTGAAGGAATCTGAGAGATATCATATGCAAAACAGCAATCTCCAGAGCAGAGTCCCTGGAG 2133
DB 481 CTGAAGGAATCTGAGAGATATCATATGCAAAACAGCAATCTCCAGAGCAGAGTCCCTGGAG 540
QY 2134 GCGAGCGCATGGAACGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2193
DB 541 GCGAGCGCATGGAACGAGAAAG 600
QY 2194 GAAGAGCGTCAAG 2253
DB 601 GAAGAGCGTCAAG 660
QY 2254 GAGGAGCAGCCACGCGCCCGGAAACCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2313
DB 661 GAGGAGCAGCCACGCGCCCGGAAACCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 2314 GTCCAG 2318
DB 721 GTCCAG 725

RESULT 9
BM950476 732 bp mRNA linear EST 14-MAR-2002
LOCUS
DEFINITION
UT-M-EHOP-buu-o-05-0-UT.r1 NIH_BMAP_EHOP Mus musculus cDNA clone
IMAGE:5687260 5', mRNA sequence.
ACCESSION
BM950476
VERSION
BM950476.1 GI:19434066
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 732)
NIH-MGC <http://mgi.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pyx-5.
Location/Qualifiers
1. 732
/organism="Mus musculus"

Db 478 GTTGCAATGTCTGTGACACCACTGCGCCGCTCTGCTCCAGAAATACCTCTCTCC 537
 Qy 1180 TTGGAAGAGTTTCCTCGGAGATGTCCTGATTAAGTCTTCTGTGATGAG 1239
 Db 538 TTGGAAGAGTTTCCTCGGAGATGTCCTGATTAAGTCTTCTGTGATGAG 597
 Qy 1240 AGGCTGCTGAGAGCGCTGCTGAGAGATGAGCAGCAGAGAAAGAACTGCTGTG 1229
 Db 598 AGGCTGCTGAGAGCGCTGCTGAGAGATGAGCAGCAGAGAAAGAACTGCTGTG 657
 Qy 1300 ACATTGTAAGATGAAGACGGAGAACTTCAGACGAGCAGTGTGAGCTGAGAGCGC 1359
 Db 658 ACATTGTAAGATGAAGACGGAGAACTTCAGACGAGCAGTGTGAGCTGAGAGCGC 717
 Qy 1360 CGCCA 1364
 Db 718 CGCAA 722

RESULT 12
 CBS21237 706 bp mRNA linear EST 28-MAR-2003
 LOCUS UI-M-GH0-cem-g-10-0-UI.r1 NIH_BMAP_GH0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6841475 5', mRNA sequence.
 ACCESSION CBS21237
 VERSION CBS21237.1 GI:29354592
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 706)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-rc@mail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mouse1.html>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 The following repetitive elements were found in this cDNA
 sequence: 220-337, >(GGA)n#Simple_repeat
 Seq primer: pYX-5.

FEATURES
 source Location/Qualifiers
 1..706
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6841475"
 /tissue_type="whole brain"
 /dev_stage="1, 5, and 15 days newborn"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH BMAP GH0"
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I;
 Site 2: Not I; The library was constructed according
 to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into pYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is CGAAGTGAAT. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemn Chin, Ph.D.,

BASE COUNT 219 a 143 c 247 g 96 t 1 others
 ORIGIN
 Query Match 13.8%; Score 701.8; DB 14; Length 706;
 Best Local Similarity 99.6%; Pred. No. 5,4e-143;
 Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

program coordinator."

Qy 1204 GGGATGCCCTCATTAAGCTTCTTCTGTGTGATTCAGAGGCTGCTGAGAGCCGTCCTCA 1263
 Db 1 GGGATGCCCTCATTAAGCTTCTTCTGTGTGATTCAGAGGCTGCTGAGAGCCGTCCTCA 60
 Qy 1264 GAGGATGAGCAGCAGCCAGAGAAACTGCTGTGACATTTGAAGATAAGAACGGGAG 1323
 Db 61 GAGGATGAGCAGCAGCCAGAGAAACTGCTGTGACATTTGAAGATAAGAACGGGAG 120
 Qy 1324 AACTTGAGCGAGGAGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 1383
 Db 121 AACTTGAGCGAGGAGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGCAGCAG 180
 Qy 1384 CGCAAGAGCAGAGCCGCTGCTCAGCTGAGAGCGCCGCAAGAGCAGAGAGAGAGAGCGG 1443
 Db 181 CGCAAGAGCAGAGCCGCTGCTCAGCTGAGAGCGCCGCAAGAGCAGAGAGAGAGAGCGG 240
 Qy 1444 GAGCGCCAGAGCAGAGCCGCAAGCGCAGCTGAGAGAGCAGTGTGAGAGAGCAG 1503
 Db 241 GAGCGCCAGAGCAGAGCCGCAAGCGCAGCTGAGAGAGCAGTGTGAGAGAGCAG 300
 Qy 1504 CGGAGCTGAGAGCGCAGCAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCC 1563
 Db 301 CGGAGCTGAGAGCGCAGCAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCC 360
 Qy 1564 GCAGAAAGCGAGATCTGAGAAAGCAGCAGCACTTGAATGGAGACCGGAGCAGAGAA 1623
 Db 361 GCAGAAAGCGAGATCTGAGAAAGCAGCAGCACTTGAATGGAGACCGGAGCAGAGAA 420
 Qy 1624 CTCCTGATCAGAGAGCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1683
 Db 421 CTCCTGATCAGAGAGCAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 Qy 1684 ACTCTGAGATTGATGATGAGAGCTCTGAATGACAAAGAGATCAGTGAAGAGAACTT 1743
 Db 481 ACTCTGAGATTGATGATGAGAGCTCTGAATGACAAAGAGATCAGTGAAGAGAACTT 540
 Qy 1744 CAGATATCAGGTCTGATCTGAGCAAGCCAGAGGCAAGAAATTGAGACAGCAAGAACTT 1803
 Db 541 CAGATATCAGGTCTGATCTGAGCAAGCCAGAGGCAAGAAATTGAGACAGCAAGAACTT 600
 Qy 1804 AGAGAGCTAAGAAATTGCTGAATCAGCCACTTACAGCAGCAGTGTGAGAACTTCAGCAA 1863
 Db 601 AGAGAGCTAAGAAATTGCTGAATCAGCCACTTACAGCAGCAGTGTGAGAACTTCAGCAN 660
 Qy 1864 ATGCTTGAAGAGCTTATTCAGAGAAACAGATACAGTACAGT 1909
 Db 661 ATGCTTGAAGAGCTTATTCAGAGAAACAGATACAGTACAGT 706

RESULT 13
 CBS21154 703 bp mRNA linear EST 28-MAR-2003
 LOCUS UI-M-FY0-cfh-m-16-0-UI.r1 NIH_BMAP_FY0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6649689 5', mRNA sequence.
 ACCESSION CBS21154
 VERSION CBS21154.1 GI:29360627
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 703)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished

```

Db      1  CCAAGTGAAGAGCCAGAGATGGCTTGGAGAGAGCTGAAAGGAGACGGAGATGTTCC 60
Qy      2628 TGCAGAACTATGAGAGAGAGATTCAGAGAAATAGAGTTCCACTCCAGCCAAACAGTAGAC 2687
Db      61  TGCAGAACTATGAGAGAGAGATTCAGAGAAATAGAGTTCCACTCCAGCCAAACAGTAGAC 120
Qy      2688 CGATCTGACATCTGCCCCCTGCCCCCAAACTGGCTCTGCGTAGAACCCCTGCTCTTGGC 2747
Db      121  CGATCTGACATCTGCCCCCTGCCCCCAAACTGGCTCTGCGTAGAACCCCTGCTCTTGGC 180
Qy      2748 AGTACCTCTTCTGAGAGCCCTCCACACACCCCAACAACTGGGCGAGACTTCAGTTCACGTG 2807
Db      181  AGTACCTCTTCTGAGAGCCCTCCACACACCCCAACAACTGGGCGAGACTTCAGTTCACGTG 240
Qy      2808 GCCAGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2867
Db      241  GCCAGAGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy      2868 TTCTTGACCGTAAGTCTGAGTCTGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2927
Db      301  TTCTTGACCGTAAGTCTGAGTCTGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Qy      2928 AGCAGTCTGCTCTCCCTCCATCTCCCTGCTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAG 2987
Db      361  AGCAGTCTGCTCTCCCTCCATCTCCCTGCTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      2988 AGGCAAGAGCCCTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3047
Db      421  AGGCAAGAGCCCTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      3048 TGAAGTATCAGCCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3107
Db      481  TGAAGTATCAGCCGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      3108 GAAAGGTTGGTTCCTCCCAAGTCTTAACTGAAATCTATTTCAGGCGCCGTAAAGAAATCCAC 3167
Db      541  GAAAGGTTGGTTCCTCCCAAGTCTTAACTGAAATCTATTTCAGGCGCCGTAAAGAAATCCAC 600
Qy      3168 AACCATCGATCTGAGGCTTAAAGTCTGCTAGCTTAAAGAGTGGCTTCCCGGCG 3227
Db      601  AACCATCGATCTGAGGCTTAAAGTCTGCTAGCTTAAAGAGTGGCTTCCCGGCG 660
Qy      3228 CGCAGAGCCAGCCATTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3287
Db      661  CGCAGAGCCAGCCATTCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Qy      3288 GCAGAGAGATTAACTT 3305
Db      720  GCAGAGAGATTAACTT 737

RESULT 11
LOCUS   BU704308                729 bp      mRNA      linear      EST 09-OCT-2002
DEFINITION  UI-M-FOO-bzs-j-21-0-UI.r1 NIH_BMAP_FOO Mus musculus cDNA clone
ACCESSION  BU704308
VERSION    BU704308.1 GI:23632308
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM  Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 729)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

```

```

FEATURES
source
    Clone Distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    This clone was contributed by the Brain Molecular Anatomy Project
    (BMAP)
    Seq primer: pYX-5.
    Location/Qualifiers
        1..729
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE: 6406772"
            /tissue_type="whole brain"
            /dev_stage="embryo 12.5dpc"
            /lab_note="DH10B (T1 phase resistant)"
            /clone_11b="NIH_BMAP_F00"
            /note="Organ: Brain; Vector: pYX-Arc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Arc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is TAGAGAGCC. This library was created for the University
            Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
            program coordinator."
BASE COUNT  176 a      206 c      174 g      172 t      1 others
ORIGIN
Query Match      13.9%; Score 704.8; DB 13; Length 729;
Best Local Similarity 99.2%; Pred. No. 1.2e-143;
Matches 719; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
Qy      640  GCTGTGCTCTCTGTCGCAATGGGCTTCATTCAGTGTGGAATGCTTCACCTTAACTA 699
Db      1  GCTGTGCTCTCTGTCGCAATGGGCTTCATTCAGTGTGGAATGCTTCACCTTAACTA 60
Qy      700  TCTTCGTCCTCCAGAGAGAGTGCCTCCGCGTAAAGGGGCTCCCGCTCATACAG 759
Db      61  TCTTCGTCCTCCAGAGAGAGTGCCTCCGCGTAAAGGGGCTTCCTCGTCAACAG 120
Qy      760  CCTTGCCCTGCGTTCGCCATCTGCAAGCAGCATGGCCAAAGAGTTCCTTCAGCAGA 819
Db      121  CCTTGCCCTGCGTTCGCCATCTGCAAGCAGCATGGCCAAAGAGTTCCTTCAGCAGA 177
Qy      820  TCTGTCAGAGGTCAATTAACACTAAGTTACAGAGAGCAATCAATTCATGTCGCC 879
Db      178  TCTGTCAGAGGTCAATTAACACTAAGTTACAGAGAGCAATCAATTCATGTCGCC 237
Qy      880  AGCGCCCTCCAGAGAGAGAGTGGCTGCTCATGATCAAGGCTGAATATACAGGAG 939
Db      238  AGCGCCCTCCAGAGAGAGAGTGGCTGCTCATGATCAAGGCTGAATATACAGGAG 297
Qy      940  TTATTCACAGCAGCAGCAAAACTATGAGTGAGACATTAAAGGTCCCGAGCAGAACT 999
Db      298  TTATTCACAGCAGCAGCAAAACTATGAGTGAGACATTAAAGGTCCCGAGCAGAACT 357
Qy      1000  ATTTCATGCAATCAAGTTTAACTCCAGGCTCAGCTGCTTCATATGGAATCTTCTGAC 1059
Db      358  ATTTCATGCAATCAAGTTTAACTCCAGGCTCAGCTGCTTCATATGGAATCTTCTGAC 417
Qy      1060  ATTGATCAAGATGGAAGAACTCACTGCAAGAAATTAATCCATGATACCTAATTAAT 1119
Db      418  ATTGATCAAGATGGAAGAACTCACTGCAAGAAATTAATCCATGATACCTAATTAAT 477
Qy      1120  GTTGCATGTCGTCAGCAGCACTGCGCGCTGCTGCTCAGATATACCTCTCTCC 1179

```

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES

source

Seq primer: PYX-5.

Location/Qualifiers

1..703

/organism="Mus musculus"

/strain="CS7BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6849689"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17,5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FY0"

/note="Organ: Brain; Vector: PYX-Asc; Site: 1: Ecor I; Site: 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGCAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 228 a 154 c 225 g 94 t 2 others
 ORIGIN

Query Match 13.8%; Score 699.4; DB 14; Length 703;
 Best Local Similarity 99.6%; Pred. No. 1.8e-142;
 Matches 700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1957 CTCAAAAGGCTTGGAGCAAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCTGGAC 2016
 1 CTCAAAAGGCTTGGAGCAAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCTGGAC 60
 2017 GAGGTGGAGAGAGAGCAGCTCAAGCTGCAGAGATTGTTTCAACCAACAGCTG 2076
 61 GAGGTGGAGAGAGAGCAGCTCAAGCTGCAGAGATTGTTTCAACCAACAGCTG 120
 2077 AAGGAAGTGGAGAGATCATATGCAAAAGCAAACTCCAGAGCAGAGTCCCTGGAGGCA 2136
 121 AAGGAAGTGGAGAGATCATATGCAAAAGCAAACTCCAGAGCAGAGTCCCTGGAGGCA 180
 2137 GCGGCACTGAACAGAAAGAGAGAGAGAGAGCTGGAGTTGAGAGAAAGAA 2196
 181 GCGGCACTGAACAGAAAGAGAGAGAGAGAGCTGGAGTTGAGAGAAAGAA 240
 2197 GAGGCTCAGAGAGAGATTGAGAAAGGAGCAAGCATGGCTGGAGCATGGAGAGAG 2256
 241 GAGGCTCAGAGAGAGATTGAGAAAGGAGCAAGCATGGCTGGAGCATGGAGAGAG 300
 2257 GAGGAGCCACGCCCCCGGAAACCCCAAGAGAGAGAGAGCTGAAGAGGAGAGAGAGTGC 2316
 301 GAGGAGCCACGCCCCCGGAAACCCCAAGAGAGAGAGAGCTGAAGAGGAGAGAGTGC 360
 2317 AGGAAGTGC 2376
 361 AGGAAGTGC 420
 2377 TTCATCCGCATCAGGAGCAGCTAAGCTGGGCAACCCAGAGCAACCTGGCTTACCAAGAG 2436

Db

421 TTCATCCGCATCAGGAGCAGCTAAGCTGGGCAACCCAGAGCAACCTGGCTTACCAAGAG 480

Qy

2437 AAAGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTGTATATTACCGAGCGCTG 2496

Db

481 AAAGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTGTATATTACCGAGCGCTG 540

Qy

2497 TACCCCTTGAATCCAGAGTGCAGATGATGATACCATTCAGCCAGAGAGATTTACTCATG 2556

Db

541 TACCCCTTGAATCCAGAGTGCAGATGATGATACCATTCAGCCAGAGAGATTTACTCATG 600

Qy

2557 GTGATGAAGCAGACCTGAGAGAGCAGAGATGCTTGGAGAGAGACTGAAAGAGAGAGC 2616

Db

601 GTGATGAAGCAGACCTGAGAGAGCAGAGATGCTTGGAGAGAGACTGAAAGAGAGAGC 660

Qy

2617 GGATGTTCCCTGCAAACTATGCAAGAAAGATTCAGAAATG 2659

Db

661 GGATGTTCCCTGCAAACTATGCAAGAAAGATTCAGAAATG 703

RESULT 14

CA329050

LOCUS

CA329050 713 bp mRNA linear EST 27-NOV-2002

DEFINITION

UI-M-FY0-cdc-1-23-0-UI_r1 NIH_BMAP_FY0 Mus musculus cDNA clone

ACCESSION

CA329050.1 GI:24547148

VERSION

CA329050.1 GI:24547148

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

AUTHORS

1 (bases 1 to 713)

TITLE

NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/BLM at:
<http://imgc.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

The following repetitive elements were found in this cDNA

Sequence: 341-458, >(GGA)n\$imple_repeat

Seq primer: PYX-5.

FEATURES

source

1..713

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CS7BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6827736"

/tissue_type="whole brain"

/dev_stage="embryo 13.5,14.5,16.5,17,5dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH_BMAP_FY0"

/note="Organ: Brain; Vector: PYX-Asc; Site: 1: Ecor I; Site: 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGCAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the

Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
Program Coordinator."

BASE COUNT 199 a 157 c 245 g 106 t 6 others

ORIGIN

Query Match 13.6%; Score 691.2; DB 14; Length 713;
Best Local Similarity 98.6%; Pred. No. 1.1e-140;
Matches 704; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 1083 TGCAGAAAGATTATTCCTAGCTATGACCTAATGATGTCCTGCTGTCAGCCACT 1142
Db 1 TGCAGAAAGATTATTCCTAGCTATGACCTAATGATGTCCTGCTGTCAGCCACT 60
QY 1143 GCCGCCGCTCTGCTCTCCAGAAATACATCCCTCTTCCCTTCCAGAAAGTTCCTCCGCGAG 1202
Db 61 GCCGCCGCTCTGCTCTCCAGAAATACATCCCTCTTCCCTTCCAGAAAGTTCCTCCGCGAG 120
QY 1203 TGGGATGTCCTGATTAAGCTCTTCTTCTGATGATCAGAGGCTGCTGAGAGCCGCTGTC 1262
Db 121 TGGGATGTCCTGATTAAGCTCTTCTTCTGATGATCAGAGGCTGCTGAGAGCCGCTGTC 180
QY 1263 AGAGATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1322
Db 181 AGAGATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240
QY 1323 GAACCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1382
Db 241 GAACCTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
QY 1383 GCGCAAAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1442
Db 301 GCGCAAAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
QY 1443 GGAAGCCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1502
Db 361 GGAAGCCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
QY 1503 GCGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1562
Db 421 GCGGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 1563 CGCAAAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1622
Db 481 CGCAAAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
QY 1623 ACTCTGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1682
Db 541 ACTCTGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
QY 1683 GACTCTGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1742
Db 601 GACTCTGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 659
QY 1743 TCAGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1796
Db 660 TCAGATATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 713

RESULT 15
BG922629 782 bp mRNA linear EST 05-JUN-2001
LOCUS BG922629
DEFINITION 602821583f1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:4950524 5',
mRNA sequence.

ACCESSION BG922629
VERSION BG922629.1 GI:14303105
KEYWORDS EST.
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 782).
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

6

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contract: Robert Strauberg, Ph.D.
Email: cgabs-femail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
Plate: LLM10905 row: e column: 21
High quality sequence stop: 716.
Location/Qualifiers

FEATURES
SOURCE

1. 782
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4950524"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT;
Library constructed by Life Technologies. Investigator
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 224 a 187 c 191 g 180 t

ORIGIN

Query Match 13.6%; Score 691.2; DB 12; Length 782;
Best Local Similarity 97.4%; Pred. No. 1.2e-140;
Matches 746; Conservative 0; Mismatches 13; Indels 7; Gaps 4;
QY 3567 ATGTGGGAAGAGAACTGCAAGCTCGAGGGAAGAAACGCGAGATAGGATGTTCCAGC 3626
Db 1 ATGTGGGAAGAGAACTGCAAGCTCGAGGGAAGAAACGCGAGATAGGATGTTCCAGC 60
QY 3527 -AAATTATGTCAACTTCTAAGCCCGGAAACAAGCAAAATACCCCAATGAGTACCA 3665
Db 61 GAAATATGTCAACTTCTAAGCCCGGAAACAAGCAAAATACCCCAATGAGTACCA 120
QY 3686 AGACCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3745
Db 121 AGACCGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
QY 3746 ACGATGACGAATGACCTTCAGCAAGGCGCATCATCAAGTCTCTCAACAAAGAGAGC 3805
Db 181 ACGATGACGAATGACCTTCAGCAAGGCGCATCATCAAGTCTCTCAACAAAGAGAGC 240
QY 3806 CGGACTGGTGAAGAGAAAGTCAAGTGGGCAAGTGGGCTTCCATCAATATATGTA 3865
Db 241 CGGACTGGTGAAGAGAAAGTCAAGTGGGCAAGTGGGCTTCCATCAATATATGTA 300
QY 3866 AGCTGACCAAGATAGTACCCAGCCAGCCAGCAATGAATATATGTTGCAATCCCCCTC 3925
Db 301 AGCTGACCAAGATAGTACCCAGCCAGCCAGCAATGAATATATGTTGCAATCCCCCTC 360
QY 3926 AGGCTGAAGTCTCAAGAGACCACTATCCATATCACTGCGCCAGAGGATGATGG 3985
Db 361 AGGCTGAAGTCTCAAGAGACCACTATCCATATCACTGCGCCAGAGGATGATGG 420
QY 3986 AGATGACGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4045
Db 421 AGATGACGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480
QY 4046 ACTCACTGACAGAGATTTACTTATTTGACCTTATTTGATGATGATGATGATGATGATG 4105
Db 481 ACTCACTGACAGAGATTTACTTATTTGACCTTATTTGATGATGATGATGATGATGATG 540
QY 4106 TCACCTGGGTGACAGGAGCAAAATTTGACAACTGACAGGATGATGATGATGATGATG 4165

```

Db      541  |||||
          TCACTGCGTCAGAGCGAGAAATGCAAGACTGCACAGGAGTGGTCTTTGG 600
QY      4166 GGGTTTCCCTAGTCACTCAGACTGACCGGCCCCGCTTCAACAGGGGCTTCAATAGTT 4225
Db      601  GGGTTTCCCTAGTCACTCAGACTGACGG--CCGCTTCACACGGGCGCTTCAATAGTT 657
QY      4226 TAAGATTATTTTAAATGTATTTTACCTTTTAAATPAAAAATCTCAATCAATTACTTCT 4285
Db      658  TAAGATTATTTTAAATGTATTTTACCTTTTAAATPAAAAATCAATCAATTACTTCT 716
QY      4286 TTGCTATTTTGGTTTAC--AAAAACACCCACTATCAAGAGTGC 4329
Db      717  TTGCGTATTATGTTTACCAAAAAAGACCACTATCAAGAGTGC 762

```

Search completed: December 5, 2003, 00:39:52
 Job time : 6505.62 secs

THIS PAGE BLANK (USPTO)


```

Db      369  COACTAATCTGCTGTTTGGGATGGGAGCATGCCCAATCTTCATTCATGAGCA 428
Qy      634  CTCACAGCTGTGTCTCTGTGCCA-----TGAGCTCAATT 669
Db      429  TTGCTTCAGTTGCACTATAGCAACCCCTGTCTCTGCTACTTCAGGGCCAGTATTT 488
Qy      670  CGAGTTGTGGAATGTCTCAACCCCTTAGATCTGTGCTCCCGCAGCAGAGTGGCTCC 729
Db      489  CTTCCCTTAATGATCTGCTGCTCTAGTGCCTCTCTGTTAGTACATCTCATTAACAA 548
Qy      730  CTGGCTAAAGGGCTCTCCCGCTCATAGACCTGTGCTGTGGCCTTCCTGACGCC 789
Db      549  GGAATGCGCAGTCTATTGAGCTTTATTCATCTCTTATCTTCAACATGCTCAT 608
Qy      790  ACATGGCCAAAGATTTCTTCTTCAAGATTTGGTCCAGGGTCAAAATTAACACTAAG 849
Db      609  GCATCATTTACAGCTGATGATGGAGGATTTGGTGGTGTAGTATCCAGAGGCCGAG 668
Qy      850  TTACGAAGGCAATCATTTGATGTGGCAGCGCCCTCCAG----- 892
Db      669  TCTGTATTGATTTAGATCTAGTAGCTCAACTTCTCACTGCTTCCCTCTCAGGGAC 728
Qy      893  -----CAGCAGAAATGGCTGTGCTCTCAGTCAATCAAGGCTGAATATACAG 936
Db      729  TCACCTAAGACAGGAGCTCAGAGTGGGAGTTCTCAGCCCTCAAGATTAAGTATCG 788
Qy      937  CAGTTATTCAACAGCCAGCAAAATCTATGATGACACTTAAGTCTCCAGGCAAGA 996
Db      789  CAAAATTTAATATGCTTAGACAAAGCATGAGCGATACCTCTCAGTTTTCAACTAGA 848
Qy      997  ACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTCAATATGAAATTTCT 1056
Db      849  AATGCCCTTCTTCACTCAAAATCTCTCAAACTCAGTAGTACTATTTGAGCTTGCT 908
Qy      1057  GACATTTGATCAAGATGAGAAATCTCAGCAGAAATTTATCTTAAGTATGACCTAAT 1116
Db      909  GACATCGATGGTGAAGGACAGTGAAGCTGAAGATTTATCTGGCATGACCTCACT 968
Qy      1117  GATGTGCCATGTCTGTGACCACTGCCCGCTCTGCTCTCAAGAAATACATCCCTCT 1176
Db      969  GACATGGCCAAAGCTTGACAGCCACTACACTGACGTTGCCCTCCGAGCTGTCTCTCA 1028
Qy      1177  TCCTTCAGAAAGTTCGCTCCGGCAGTGGAAATGTCCGTCATTAAGTCTTCTTCTGTG 1236
Db      1029  TCTTTCAGAGG-----GGAAAGCAAGTTGATTCGTAT 1064
Qy      1237  CAGAGGCTGCTGAGAGCCGCTCTCAGAGATGAGCAGCCAGAGAAATCTGCT 1296
Db      1065  GGAATCTGCTTCATATCAAGAAACACAGAG--AAGAGCTCAGAGAAAGTGGCA 1121
Qy      1297  GTGACATTTGAAGATTAAGAGCGGAGAACTTCAGCCAGAGGAGTGTGAGTGAAG 1356
Db      1122  GTTACTTTTGAAGCAAAACGAAAGCCAACTATGAACAGAGAAATGAGTGAAGAG 1181
Qy      1357  CGCGCCCAAGGCTCTTGAAGAGAGCGCAAAAGAGCAGAGCGTTGGCTCAGCTGAG 1416
Db      1182  CGAGCCCAAGTGTATGAGAGAGAGAGGAGGCTGAACGAAAGCCAGAAAGAG 1241
Qy      1417  CGCGCCGAGAGAGAGAGAGAGCGGAGCGCCAGAGAGAGAGAGCGCAAGCGAGCTG 1476
Db      1242  AAGGAAGAGTGGAGCGGAAACAGAGAACTGCAAGAGCAGAAATGGAAGAGAGCTG 1301
Qy      1477  GAGCTGAGAGAGAGAGAGAGAGCGGAGAGCTGAGCGGACGAGAGAGAGAGAG 1536
Db      1302  GAGTTGAGAGAAACGCTTGAAGAAACAGAGAGCTGAGAGAGCAGCGGAGAGAGAG 1361
Qy      1537  AGGAAGAGATTCAGAGAGCGGAGCGGAGAACTGGAAGAGCAGAGCAACTT 1596
Db      1362  AGAAAGAGATGAAGAGAGAGAGCGGAGCGGAGCTTGAAGAGCAAGCGCTTTA 1421
Qy      1597  GAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1656

```

```

Db      1422  GAATGGAGAGAGATCCGTCGAGAGAGCTGCTCAGTCAGAGAGACCGAGGAAAGAGAC 1481
Qy      1657  ACCGTGCTCTGAAGAGAGAGAGAGAGTCTGAGAGTTGAGTTAGAACTCTGAATAC 1716
Db      1482  ATTGTCAAGCTGACCTCCAGAAAGAAAGTTCCACCTGGAACGTGAAGCAGTAAATGA 1541
Qy      1717  AAAAGCATCAGCTAGAGAGAAACTTCAGAGATTCAGAGTTCGACTGCGCAACCAGAG 1776
Db      1542  AAACATCAGAGATCTCAGAGAGACTACAGATGTCCAAATCAGAAAGCAACAAAG 1601
Qy      1777  CAAGAAATTGAGACAGCAACAGTCTAGAGAGCTTAAGAAATTTGCTGAATACCCACTTA 1836
Db      1602  ACTAGCTAGAGAGTTTGGATTAACAGTGTGACCTGGAATTTAGAAATCAACACTT 1661
Qy      1837  CAGCAGAGTTGAGAGATCTCAGCAAAATGTTGGAAGACTTATTCAGAGAAACAGTA 1896
Db      1662  CAACAGAGCTTAAGAAATCAAAATTAAGTTATCTATCTGGTCCCTGAGAAACAGCTA 1721
Qy      1897  CTCAGTACAGTTAAACAAAGTCCAGCAGAACAGTTTGCATAGAGACTCGCTTTPACC 1956
Db      1722  TTAAACGAAAGAAATTTAAACATGACCTCAGTAAACACACTGATTCAGGATCAGTTTA 1781
Qy      1957  CTCAAAAGAGCTTGAAGAGAGAGAGAGCTGCCGAGCAGCTCCGGAGCAGCTGAC 2016
Db      1782  CTTCATAAAGATCATCAGAAAGAGAAAGATTATGCCAAAGACTTAAAGAACTTAAGAT 1841
Qy      2017  GAGGTGAGAGAGAGACAGGTCAAGCTGAGAGAGATGATGTTTCAACAAACAGCTG 2076
Db      1842  GCTCTGAAAAGAACTGACATCTAAGCTCTCAAAATGATTTCAATTAACATCAGCTG 1901
Qy      2077  AAGGATGAGAGATACATATAGCAAAACAGCACTCCAGAGCAGAGGTCCCTGAGGCA 2136
Db      1902  AAGGATCAGAGAAAGCTATATACACAGAGATTAGCCCTTGAACAACTTCAATTAATC 1961
Qy      2137  GCGGCACTGAAGCAGAAAGAGCAGAGAGAGAGAGAGCTGAGTTAGAGAGCAAAA 2192
Db      1962  AACGTGACAAATTTGAAGAAATCGAAAGAAAGATTAGCAAAAAA 2017

```

```

RESULT 2
US-09-338-933-72
; Sequence 72, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338, 933
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-72

```

```

Query Match      9.5%; Score 484.8; DB 4; Length 2017;
Best Local Similarity 55.4%; Pred. No. 4,36-116;
Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

Qy      226  GGCCTGATCAGCAAGGTGAAGTAATGAACCATGCTCATGTTTCCACACCTTTGGGT 285
Db      9    GGTGAGAGCTGCAAGAAAGTCAAGATCATGATGCTCAGTTTCCACAGCATGAT 68
Qy      286  GGTAGCTGATGTCTGGGCCATATCTGTGAGGAAGGCGCAAGCATGACCAAGTTC 345
Db      69  GGAAGGCCAAATATGTGGCTATTACATCTGAAGAGATTAAGCATGATTAACAAGTTT 128
Qy      346  CTTAGCTGAAGCGGATAGCGGATTTATTACTGTGATCAAGCGAGAACTTTTTC 405

```

Db 129 GATTAACCTCAACCTTCAGAGGTTACATAACAGTGATCAAGCCCTACTTTTCTTA 188
 Qy 406 CAATCTGGTTTACCTGACCTGTCTTAGCAATAATATGGGCGGTAGGGACATGAATAAC 465
 Db 189 CATCTAGGTTGCGGGCCCGGTTTATGTAATAATATGGGCTTATCAATGATCGAAGAC 248
 Qy 466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTATCAATGAAGCTA 525
 Db 249 GATGGGAAGATGACCAAGAGTTCTCTATAGCTATGAATCACTCAATGAATGAAGT 308
 Qy 526 CAAGATATAGCTCCCTCCCACTTCCCTCTGTCAATGAAGCAAGCAAGT--GCT 582
 Db 309 CAGGGCCCAAGCTGCTGTAGTCTCCCTCTATCAATGAAGCAAGCCCTATGTTCTCT 368
 Qy 583 ATTTCAGTGCACAGATTTGTAATAGAGGATTTGTAAGT--CCACCA 633
 Db 369 CCACTAATCTGTGCTGTGGATGGGAAGATGCCAATCTGTCAATTCATCAAGCA 428
 Qy 634 CTCACAGCTGTGCTGCTGTGCA--TGGGCTCATT 669
 Db 429 TTGCTCCAGTTGACCTATAGCAACCTGTCTGTCTACTTCAAGGACCAATTT 488
 Qy 670 CCAATGTTGGAATGTCTCACCCTTAGATTTTCTGTCTCCCTCAAGAGAGTCCCTCC 729
 Db 489 CTTCCCTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548
 Qy 730 CTGGCTAAGGGGCTCCTCCCTCATACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
 Db 549 GGAATGCAAGTCTCAATTCAGCTTATTCATTCCTTATTTCTTCAACATTCCTCAT 608
 Qy 790 ACATGAGCAAAAGATTTCTCTTCAAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 849
 Db 609 GATCATCTTAACAGCTGATGATGAGAGATTTGCTGCTGCTGCTGCTGCTGCTGCTG 668
 Qy 850 TTAAGAGGCAATCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 892
 Db 669 TCTCTATGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAGATTTAG 728
 Qy 893 -----CAGAGAAATGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 936
 Db 729 TCACTTAAGACAGGACCTTAGAGTGGGAGTTCCTAGGCTTCAAGATTTAAAGTTCGG 788
 Qy 937 CAGTTATTCACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 996
 Db 789 CAAAATTTAATGTTCTAGACAAAGCATAGAGGATCTCTCAGGTTTCAAGCTGAG 848
 Qy 997 ACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCTTCAATATGAAATCTTCT 1056
 Db 849 AATGCCCTTCTCAGTCAATCTCTCTCAAACTCAGTAGCTACTATTGGAATCTGGCT 988
 Qy 1057 GACATTCATCAAGATGGAATCTCATCTGCAAGAAATTTTCTAGCTATGACCTTAAT 1116
 Db 909 GACATGATGATGACGACGAGTGAAGCTGAAGAAATTTATTCGGGATGACCTCCTAC 968
 Qy 1117 GATGTTGCAATGTCTGTCAGCAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1176
 Db 969 GACATGCGCAAGCTGACGACGACCTACCTACCTGCTGCTGCTGCTGCTGCTGCTG 1028
 Qy 1177 TCTCTCAGAAAGTGTGCTCGGAGTGGATGCTGCTCAATGCTTCTTCTGTGAT 1236
 Db 1029 TCTTTTCAAGG-----GGAAAGCAAGTGAATCTGTTAAT 1064
 Qy 1237 CAGAGGCTGCTGAGAGAGCTGTGCTGAGAGATGACAGAGCAAGAAAGAAATCTGCT 1296
 Db 1065 GGAATCTGTCTCATATCAAGAAACACAAAG--MAGAGCTCAGAAAGAACTGCCA 1121
 Qy 1297 GTGACATTTGAAGATTAAGAGCGGAACTTGAAGAGGAGCTGTGAGCTGAGAG 1356
 Db 1122 GTTACTTTTGAAGCAAAACGAAAGCACTATGAAAGAAACATGAGCTGAGAGAG 1181
 Qy 1357 CGCGCCAAAGCGCTCTTGAAGCAAGCGCAAGAGAGAGAGGCTGTGCTGAGCTGAG 1416
 Db 1182 CGAGCGCAAGTGTGATGAGAGCAGAGAGGAGGAGGCTGAACGCAAGGCCAGAAAGAG 1241

Qy 1417 CGCGCCGAGAGAGAGAGAAAGAGCGGAGCCGACAGAGAGAGGCAAGCGGACCTG 1476
 Db 1242 AAGAAAGAGTGGAGGAGGAAACAGAGAGATCTGCAAGAGCAAAATGAGAAAGAGCTG 1301
 Qy 1477 GAGCTGAGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
 Db 1302 GAGTTGAGAGAGAGAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
 Qy 1537 AGAAGAGATCGAG 1596
 Db 1362 AGAAGAGATGAG 1421
 Qy 1597 GATGAG 1656
 Db 1422 GAATGGAG 1481
 Qy 1657 ACCGAGTCTGAG 1716
 Db 1482 ATTGTAGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1541
 Qy 1717 AAAAGCATGAGCTGAG 1776
 Db 1542 AAACATCAGAGAGATCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1601
 Qy 1777 CAGAAATGAG 1836
 Db 1602 ACTGAGCTGAG 1661
 Qy 1837 CAGAGAGATGAG 1896
 Db 1662 CAACAG 1721
 Qy 1897 CTCAGTACAGTAAAG 1956
 Db 1722 TTAAG 1781
 Qy 1957 CTCAG 2016
 Db 1782 CTTGATTAAG 1841
 Qy 2017 GAGGTGAG 2076
 Db 1842 GCTCTGAG 1901
 Qy 2077 AAGGAG 2136
 Db 1902 AAGGAG 1961
 Qy 2137 GCGGAG 2192
 Db 1962 AAGGAG 2017

RESULT 3
 US-09-215-681-72
 ; Sequence 72, Application US/09215681A
 ; Patent No. 6528253
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Prudakis, Tony N.
 ; APPLICANT: King, Gordon B.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
 ; FILE REFERENCE: 210121.463
 ; CURRENT APPLICATION NUMBER: US/09/215,681A
 ; NUMBER OF SEQ ID NOS: 310
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 72
 ; LENGTH: 2017
 ; TYPE: DNA
 ; ORGANISM: Homo sapien


```

; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOMIKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TEXT: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 193:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2873 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-630-915A-193

```

Query Match 6.6%; Score 333.2; DB 4; Length 2873;

Best Local Similarity 61.08; Pred. No. 1.7e-76; Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

```

QY 2867 CTCTCTGACCGTACCTAGTGTGGCCAGTTACGCGAGATCAGCTTTACCCGACCA 2926
DB 541 CTTTTCMAACCTTACTTAAATACATGACGAGAAATACGCTTCACTGAACTG 600
QY 2927 CAGCCACTGGCTCTCCCATCTCCCGTCTGGGCCAGGCTGAAAAGTGAAGGCTAC 2986
DB 601 TGTCCCTGG--ATCTGATCACTATTCATGAGAGGAGCAAGGTGTAAGAACTTAA 657
QY 2987 AACCGCAAGCCCTGTATCTCTGAGAGCCCAAAAAGACAACCACTTAATTAAAGAAA 3046
DB 658 AACCAAGAGCCCTTGTCTCTGAGCTGCAAGAAAGATTAACCACTTCTCAAAAC 717
QY 3047 GTGACGTCATCACGTTCTTGAAGACAGACAGATGTGTGAGTTGAGAACTCAAGTCT 3106
DB 718 ATGACATATTAATCTGCTTGGAGACAGCAAGAAATGTTGGAGAGTCAATGAG 777
QY 3107 AGAAGGTTGGTCCCAAGTCTTACGTGAATCTCAATTTCAAGGCGCTGAAGAAATCCA 3166
DB 778 GAGAGAGATGTTTCCAAATCTTATGTCAAGATCAATCTGGAGAGTAAGTAAACGGG 837
QY 3167 CAAGCATCATCTAGGCGCTTACGAAAGTCTCTGTACTTAAGAGAGTGGCTTCCCGG 3226
DB 838 AA-----GAACCAAGAGCTTTGTATCAGCTGTAAATGAAGAACTTCACTCGG 885
QY 3227 CCGCCAAAGCAGCATTCGCGAGAGAGATTATGTCATGTACACATACAGAGTTCTG 3286

```

```

DB 886 CAGCCTATTACGT-----TGGAGAGAAATATATTCACCTTATCATATTCAGTGTGG 939
QY 3287 AGCAAGAGATTTAACTTCTTACAGCAAGGAGATGTATTTGTGTACCAAGAAAGATGCTG 3346
DB 940 AACCTGAGATTTGACTTTTCAACGAAAGTGAAGAAATATTTGGTACCCAGAAAGATGAG 999
QY 3347 ACTGTGAGACGGGAGCGTGGGCAAGATCCGGAAGTCTTCCCTTAACTATGATGAGGC 3406
DB 1000 AGTGTGACAGAGATATTTGAGATGAGATGAGAAATTTTTCATTAATATGTCAAAAC 1059
QY 3407 TTAAGATTACAGAGGCTCTGAACTGTGGGAAAACAGGAGATTGAGAAAAAAACCTG 3466
DB 1060 CAAAGATCAAGAGATTTTGGAGTCTTGAAGATCTGAGAGATCAATTAATAAAACCTG 1119
QY 3467 AATTGCCAGATTTATGCTCTTCTTACGCTGCTACTAGTGTCCGAAACACTTACCTGCTC 3526
DB 1120 AGATTGCTCAGGATTAATCTTACAGATATGTGTCTTGTGTAACAACTTACCTTGAC 1179
QY 3527 CTGGCAGCTGATTTCTGATCCGAAAGAAACCCAGGTGATGTGGAGAGATCTG 3586
DB 1180 CAGGACAGTATATTAATTAATTAAGAAATTAAGAAATGAGTGTGGCAAGAGATTCAG 1239
QY 3587 AAGCTCGAGGAGAAAAGCCGCAATAGGTGTTTCCAGCAATATGTCAAACTTCTAA 3646
DB 1240 AGGCCAGAGAGAAAAGCCAGCAAGAAAGATGTTTCTTCCAGTCAATGTAAACTTTGG 1299
QY 3647 GCGCCGAAACAGCAAAATACCCCACTGAGCTACCAAGACCGAGTGCAGCCAGCAG 3706
DB 1300 GTCCAGTATGTAAAGAGCCACACCTGCTTCATC-----CTG 1338
QY 3707 TGTCCAGATGATCGGATGATGATTAACCGCCCAAGATGCAATGATGCTTCA 3766
DB 1339 TATGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
QY 3767 GCAAGGCGAGATCATCAAGTCTTCAACAGAGAGAGCCGAGCTGTGAAAAGAGAGAG 3826
DB 1399 CCAAGGAGACACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458
QY 3827 TCAAGTGGCAAGTGGGCTCTTCCATCCATCAATTAATGTAAGCTGACAGACAGATGAC 3886
DB 1459 TCAACGGGGTGAAGTGTCTTCTTCCCTTCAACCTAAGTAAAGATGACAGACTCAGATC 1518
QY 3887 CCAAGCCAGCAATGAATCA 3904
DB 1519 CAAGTCAACAGTGACCA 1536

```

RESULT 5

US-08-630-915A-39

; Sequence 39, Application US/08630915A

; Patent No. 6309820

; GENERAL INFORMATION:

; APPLICANT: SPARKS, Andrew B.

; APPLICANT: HOFFMAN, No. 6309820h

; APPLICANT: KAY, Brian K.

; APPLICANT: FOMIKES, Dana M.

; APPLICANT: MCCONNELL, Stephen J.

; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

; NUMBER OF SEQUENCES: 227

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Miarock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-39

```

```

Query Match      5.1%; Score 258; DB 4; Length 747;
Best Local Similarity 63.9%; Pred. No. 3e-57;
Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 3247 GGAGAGAGATTATTTGCGATGACATACAGAGTCTGAGCAAGAGATTAACTTT 3306
DB 115 GGAGAGAGATTATTTGCGATGACATACATGAGTGGAACTGGAGATTGCTTTC 174

QY 3307 CAGCAAGGGGATGATGTTGTTGTTACCAAGAAAGTGTGACTGGTGGACGGGAACGGTG 3366
DB 175 ACAGAGAGTGAAGAAATATGTTGTTGACCCAGAAAGTGAAGTGTGACAGAAATATTT 234

QY 3367 GCGCAGATCGGAGTCTTCCCTTCTACTATGTGAGGCTTAAAGATTGAGGGCTCT 3426
DB 235 GGAGATTAAGTGAATTTTTCATCAATATGTCAAAACCAAGGATCAAGAGATTTT 294

QY 3427 GGAATGCTGGGAAAACAGGAGTTTAGGAAAAAACTGAAATTTGCCAGTTATTTCT 3486
DB 295 GGGAGTGTACAGAGTGTGAGCATCAATAAAAAACTGAGATTGCTCAGGTAACTTCA 354

QY 3487 TCCATCGTGTACATGTCCTCCGAAACACTCACTCCGCTGCTGGGCACTGATTCGTATC 3546
DB 355 GCATATGTTGCTTCTGTTCTGTAACACTTGAAGCTTGAACAGGACGTTAATTAATTT 414

QY 3547 CGGAAAAAGAAACCCAGTGTGATGTTGGGAAAGAGAACTGCAAGTCCAGGGGAAAAAGCGC 3606
DB 415 CTTAAAGAAAAATACAGTGTGTTGGTGGCAAGAGATTACAGGCCAGGAAAAAGCGA 474

QY 3607 CAGATAGGTTGTTTCCAGCAATTAATGTCAAACTTTCTAAGCCCGGAAACAAAGCAAAATC 3666
DB 475 CAGAAAGATGTTTCTGCGCAGTGTATTAACCTTTGGGTCCAAAGCAAGTGAAGAACCC 534

QY 3667 ACCCAACTGAGCTACCCAAAGCCGAGTGCAGCAGAGTGTGCAAGTATGCGGGAG 3726
DB 535 AACCTGCTCTTCACTCT-----GTAATGTCAGTGTATCTTAAG 573

QY 3727 TACGATTAACCGCCAGAACGATGACGAACTAGCTTTCAGCAAAAGGCGCAATCATCAAC 3786
DB 574 TATGCTATGACGAAATTAATGAATAGCTCAATTTCTCCAAAGGAACTCATTAAT 633

QY 3787 GTCTCTCAACAGAGAGACCCGAGCTGTGTGAAAGAGAAAGTCAAGTGGGCAAGTTGGGCTC 3846
DB 634 GTTATGAACAAAGATGATCTGATTTGTGGGCAAGAGAGATCAACGGGGTGACTGTCTC 693

QY 3847 TTCGATCCAAATTATTAAGTGAACACAGCATGAGACCCAGCAGCAAGCAATTA 3900
DB 694 TTTCTTCAAACTGATTAAGTGAACAGACGACTGATTCAGTCAAGTCAAGTGA 747

```

RESULT 6
US-09-404-879A-5

```

Sequence 5, Application US/09404879A
Patent No. 6468546
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-404-879A-5

```

```

Query Match      3.0%; Score 151.4; DB 4; Length 531;
Best Local Similarity 61.3%; Pred. No. 1.5e-29;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

QY 226 GCGCTCGATGACAGAGGTGAACGTAAATGAACCATGCTCAGTTTCCACACCTTTGGT 285
DB 9 GCGTGAAGCTGCAAGAAAGAGTCAAGTATGCTCAGTTTCCACAGCATGAT 68

QY 286 GGTAGCTGTGATGTTGCGCCATPACTGTGAGGAAAGGCGCAAGCATGACAGACTTC 345
DB 69 GAGAGGCGCAATATATGTTGGCTATTTACTCATGAGAACGTACTAGCATGATTAACAGTTT 128

QY 346 CTTAGCTGAAGCCGATAGGGGATTTATTTACTGTGATCATCAAGGAGAACTTTTTC 405
DB 129 GATACCTCAAACTTTAGAGGTTACATTAACAGGTATCAAGCCGTACTTTTCTTA 188

QY 406 CAATCTGGGTTACTCTGACCTGTCTTAGCACAATATGAGGCGCTAGCGGACATGATPAC 465
DB 189 CAGTCAGTGTGCGGCGCCGCTTTAGCTGAATATGGGCTTATCAGATCTGAACAG 248

QY 466 GATGAAGATGATGATCAAGTGAATTTTCCATPACCTGAAAGCTTATCAAACTGAAGCTA 525
DB 249 GATGGGAAGTGAACAGCAAGAGTCTCTATAGCTTGAAGCTCAAGTTAAAGTTG 308

QY 526 CAAGATATAGCTCCCTCCCTCACACTTCCCTGTGATGAAGCAAGCAAGTGGCTATT 585
DB 309 CAGGCGCAACAGCTGCTGTATGCTTCCCTCTTATGAAACAAACCC-----CTATG 362

QY 586 TCCAGTCAACCAAGATTGTTATGAGAGGATTTCTGATGACATGACCACTCAGAGCTGT 645
DB 363 TTTCTCCACTAATCTGCTGCTGTTTGGATGGAGATGCCCAATCTGTCCATTCAT 422

QY 646 GCTCTGTGCC 656
DB 423 CAGCCATTGCC 433

```

RESULT 7
US-09-338-933-5
Sequence 5, Application US/09338933
Patent No. 6488931
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer Lynn
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
TITLE OF INVENTION: OVARIAN CANCER
FILE REFERENCE: 210121.462C1
CURRENT APPLICATION NUMBER: US/09/338,933
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 531
TYPE: DNA

QY 619 GCTAGCATGCCACCACTGACAGCTGTGCTCTGAGCC 656
 DB 368 GGAAGCATGCCCAATCTGTCTCATTCATCAGCCATTGCC 405

RESULT 10

US-09-338-933-60
 ; Sequence 60, Application US/09338933
 ; Patent No. 6488931
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer Lynn
 ; APPLICANT: King, Gordon E.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
 ; TITLE OF INVENTION: OVARIAN CANCER
 ; FILE REFERENCE: 210121.462C1
 ; CURRENT APPLICATION NUMBER: US/09/338,933
 ; CURRENT FILING DATE: 1999-06-23
 ; NUMBER OF SEQ ID NOS: 312
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 60
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-338-933-60

Query Match 2.8%; Score 144; DB 4; Length 480;
 Best Local Similarity 62.1%; Pred. No. 1,2e-27;
 Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

QY 259 ATGGCTCAGTTTCCACACCTTTCGTTGAGTCTGATGTCGGGCAATACCTGTGAG 318
 DB 14 ATGGCTCAGTTTCCACACCTTTCGTTGAGTCTGATGTCGGGCAATACCTGTGAG 73
 QY 319 GAAAGGGCCCAAGCATGACGAGATTCCTTACCTGAAAGCCGATTAATTAAT 378
 DB 74 GAACGTACTTAAGCATGATTAAGTGAAGTGAATCACTCAAACTTCAGAGGATTAACA 133
 QY 379 GGTGATCAAGCAGAGAACTTTTTCCTCAATCTGGGTTACCTGACCTGTCTTACACA 438
 DB 134 GGTGATCAAGCAGAGAACTTTTTCCTCAATCTGGGTTACCTGACCTGTCTTACACA 193
 QY 439 ATATGGGCGCTAGCGACATGATTAAGATGAGATGATCAAGTGAATTTTCCATA 498
 DB 194 ATATGGGCGCTATGATGATGATTAAGATGAGATGAGATGATCAAGTGAATTTTCCATA 253
 QY 499 GCCATGAAGCTTATCAAACTGAGATCAAGATATCAAGTCTCCCTCCACACTTCCCT 558
 DB 254 GCTATGAAGCTTATCAAACTGAGATTAAGTGAAGTGAAGTCTCCCTGATGCTCCCT 313
 QY 559 GTCATGAAGCAGCAACCACTGCTATTTCCAGTGCACCACTTTTGTATGAGAGGATT 618
 DB 314 ATCATGAAGCAACCACTGCTATTTCCAGTGCACCACTTTTGTATGAGAGGATT 367
 QY 619 GCTAGATGCCACCACTGACAGCTGTGCTCTGAGCC 656
 DB 368 GGAAGCATGCCCAATCTGTCTCATTCATCAGCCATTGCC 405

RESULT 11

US-09-215-681-60
 ; Sequence 60, Application US/09215681A
 ; Patent No. 6528253
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Fradakis, Tony N.
 ; APPLICANT: King, Gordon E.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
 ; TITLE OF INVENTION: OF OVARIAN CANCER
 ; FILE REFERENCE: 210121.463
 ; CURRENT APPLICATION NUMBER: US/09/215,681A
 ; CURRENT FILING DATE: 1998-12-17
 ; NUMBER OF SEQ ID NOS: 310

; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 60
 ; LENGTH: 480
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 US-09-215-681-60

Query Match 2.8%; Score 144; DB 4; Length 480;
 Best Local Similarity 62.1%; Pred. No. 1,2e-27;
 Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

QY 259 ATGGCTCAGTTTCCACACCTTTCGTTGAGTCTGATGTCGGGCAATACCTGTGAG 318
 DB 14 ATGGCTCAGTTTCCACACCTTTCGTTGAGTCTGATGTCGGGCAATACCTGTGAG 73
 QY 319 GAAAGGGCCCAAGCATGACGAGATTCCTTACCTGAAAGCCGATTAATTAAT 378
 DB 74 GAACGTACTTAAGCATGATTAAGTGAAGTGAATCACTCAAACTTCAGAGGATTAACA 133
 QY 379 GGTGATCAAGCAGAGAACTTTTTCCTCAATCTGGGTTACCTGACCTGTCTTACACA 438
 DB 134 GGTGATCAAGCAGAGAACTTTTTCCTCAATCTGGGTTACCTGACCTGTCTTACACA 193
 QY 439 ATATGGGCGCTAGCGACATGATTAAGATGAGATGATCAAGTGAATTTTCCATA 498
 DB 194 ATATGGGCGCTATGATGATGATTAAGATGAGATGAGATGATCAAGTGAATTTTCCATA 253
 QY 499 GCCATGAAGCTTATCAAACTGAGATCAAGATATCAAGTCTCCCTCCACACTTCCCT 558
 DB 254 GCTATGAAGCTTATCAAACTGAGATTAAGTGAAGTGAAGTCTCCCTGATGCTCCCT 313
 QY 559 GTCATGAAGCAGCAACCACTGCTATTTCCAGTGCACCACTTTTGTATGAGAGGATT 618
 DB 314 ATCATGAAGCAACCACTGCTATTTCCAGTGCACCACTTTTGTATGAGAGGATT 367
 QY 619 GCTAGATGCCACCACTGACAGCTGTGCTCTGAGCC 656
 DB 368 GGAAGCATGCCCAATCTGTCTCATTCATCAGCCATTGCC 405

RESULT 12

US-08-728-323A-1
 ; Sequence 1, Application US/08728323A
 ; Patent No. 5948676
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Yuan
 ; APPLICANT: Bohenzky, Roy A.
 ; APPLICANT: Russo, James J.
 ; APPLICANT: Edelman, Isidore S.
 ; APPLICANT: Moore, Patrick S.
 ; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
 ; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
 ; TITLE OF INVENTION: Encoding Same And Uses Thereof
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cooper & Dunham LLP
 ; STREET: 1185 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/728,323A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: White, John P.
 ; REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 2.8%; Score 144; DB 2; Length 3489;
Best Local Similarity 45.1%; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

1233 GGATCAGAGGCTGCTGAGGAGCGCTGTCAGAGGATGACAGCAGCAGGAGGAAACT 1292
1578 GCAGAGGCCACAGCAGAGGAGCCACAGCAGAGGCCACAGCAGAGGAGCCACAGCA 1637
1293 GCCTGTGACATTTGAAGATAAGACGGGAGAACTTTCAGCGAGGAGTGAGGCTGGA 1352
1638 GCAGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1697
1353 GAAGCGCGCCCAAGCGCTCTTGAAGCAGCGCCAAAGACAGAGCGGTTGCTCACT 1412
1698 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
1413 GGAGCGCGCCGACAGGAGGAGAAAGCGGAGCCGACAGGAGGAGGAGCCAGGAGCC 1472
1758 GCGGAGGCCACAGCAGCGGAGCCACAGCAGGATGACAGCAGAGGATGAGCAGCA 1817
1473 GCTGAGCTGAGAGAGCAGCTGAGAGACAGCGGAGCTGAGCGGAGGAGGAGGAGCA 1532
1818 GCAGAGTGAAGCAGCAGAGGATGACAGCAGCAGAGTGAACAGCAGAGATGACAGCA 1877
1533 GAGGAGGAGAGATGAGAGGCGGAGCGGAGCCGAGAACTGGAAAGGAGCAGCA 1592
1878 GCAGAGTGAAGCAGCAGAGGATGAGCAGCAGGATGAGCAGCAGAGGATGAGCAGCA 1937
1593 ACTTGATGGGAAACCGAGACAGGAACTCTGATTCAGAGAAAGAGAGCAGCA 1652
1938 GCAGAGTGAAGCAGCAGAGGATGAGCAGCAGGATGAGCAGCAGAGGATGAGCAGCA 1997
1653 GGGCACCGTGTCTGAGAGCAGAGGAGAACTTGAAGTTGATGAGTGAAGCTCTGAA 1712
1998 GCAGAGTGAAGCAGCAGAGGATGAGCAGCAGAGTGAAGTGAAGTGAAGTGAAGCA 2057
1713 TGAAGAAAGCATCAGCTAGAGGAAATCTTCAAGATATCAGAGTGTGCACTGGCAACCA 1772
2058 GCAGAGTGAAGCAGCAGAGGATGAGCAGCAGGATGAGCAGAGGATGAGCAGCA 2117
1773 GAGGAGGAGAAATGAGAGCAGAAAGTATGAGAGCTTAAGATTTGCTGAATTCACCA 1832
2118 GAGGAGGAGGATGAGAGGAGGAGCAGAGTGAAGCAGCAGAGTGAAGCAGAGCAGCA 2177
1833 CTTCAGCAGCAGCTGAGAGGATCTCAGCAATGCTTGAAGACTTATTCAGAGAAAC 1892
2178 GATGAGGAGCAGCAGAGGATGAGCAGCAGCAGAGTGAAGCAGCAGAGTGAAGTGA 2237
1893 GATATCTAGTACCACTTAAACAGTCCAGCAGAAAGTTTGATAGAGACTTCGCTTCT 1952
2238 GCAGCAGCAGCAGAGTGAAGCAGAGCAGAGGAGCAGAGCAGAGGAGGAGCAGCA 2297
1953 TACCTTCAAGAGCCTTGAAGCAAGAGCTGCGCCGACAGCAGTCCGAGAGCAGCT 2012
2298 GCAGAGGATGAGAGCAGAGGAGGAGTGAAGAGATCAGAGCAGAGGATTAAGAGCA 2357
2013 GAGCAGGAGTGAAGAGAGCAGAGTCAAGCTGACAGAGATGATTTTCAACAAACA 2072

2358 GGAGCAGAGGATTAGAGAGCAGAGAGGATTAGAGAGCAGAGCAGAGGATTAGAGGA 2417
2073 GCTGAAGAACTGAGAGAGATCATRGCAGCACTCCAGAGAGCAGAGCTCCCTGCA 2132
2418 GCAGAGCAGAGGATTAGAGAGCAGAGCAGAGGATTAGAGAGCAGAGCAGAGTGA 2477
2133 GGAGCGGAGCTGAAGCAGAAAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2192
2478 GAGCAGAGCAGAGGATTAGAGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2537
2193 GGAAGCGCTCAGAGAGCAGATTCAAGAAAGGAGCAAGCAATGCTGAGCATGTGAGCA 2252
2538 GCAGAGCAGAGGAGTGAAGAGCAGAGCAGAGCAGAGGAGGAGGAGGAGGAGGAGGAG 2597
2253 GAGAGCAGCAGCGCGCCGAGAAACCCACAGAGAGCAGACAGTGAAGAGGAGAGCAG 2312
2598 GAGGAGCAAGAGCAGAGCAGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2657
2313 TGTCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2372
2658 GCAGAGAGCAGAGGATTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2717
2373 GCTTTCATCCGATCAGAGCAGCTAAGCTGCGCACCCAGG 2416
2718 GGAAGAGCAGAGCAGCAGAGGATTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2761

RESULT 13
US-09-298-568-1
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
APPLICANT: Kieft, Elliott D.
APPLICANT: Balleslae, Mary E.
APPLICANT: Kaye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
FILE REFERENCE: 16412-10001R
CURRENT APPLICATION NUMBER: US/09/298,568
EARLIER FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposis's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 2.8%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

1233 GGATCAGAGGCTGCTGAGGAGCGCTGTCAGAGGATGACAGCAGCAGGAGGAAACT 1292
1578 GCAGAGGCCACAGCAGAGGAGCCACAGCAGAGGCCACAGCAGAGGAGCCACAGCA 1637
1293 GCCTGTGACATTTGAAGATAAGACGGGAGAACTTTCAGCGAGGAGTGAGGCTGGA 1352
1638 GCAGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1697
1353 GAAGCGCGCCCAAGCGCTCTTGAAGCAGCGCCAAAGACAGAGCGGTTGCTCACT 1412
1698 GCGGAGGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
1413 GGAGCGCGCCGACAGGAGGAGAAAGCGGAGCCGACAGGAGGAGGAGGAGGAGGAGGAG 1472
1758 GCGGAGGCCACAGCAGCGGAGCCACAGCAGGATGAGCAGCAGAGGATGAGCAGCA 1817
1473 GCTGAGCTGAGAGAGCAGCTGAGAGCAGCGGAGCTGAGCGGAGGAGGAGGAGGAGGAG 1532

```

Db      1818 GCAGGATGAGCAGCAGCAGATGAGCAGAGGATGAGCAGCAGATGAGCAGCA 1877
Qy      1533 GAGGAGGAAAGGATCGAGAGCGCCGCAAAACGGGAATCTGGAAAGCAGCGCA 1592
Db      1878 GCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCA 1937
Qy      1593 ACTTGAATGGGAAACCGGAGACAGGAATCTCTGATCAGAGAAACAAGAGCAGGA 1652
Db      1938 GCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCA 1997
Qy      1653 GGGCACCCTGTCTTGAAGGCAAGAGAGAGACTCTGAGTTTGAATGAAAGCTTGA 1712
Db      1998 GCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCA 2057
Qy      1713 TGACAAAAGCTCAGCTAGAGAAATTTAGAGATATCAGGTCTCAGCTGCA 1772
Db      2058 GCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2117
Qy      1773 GAGCAGAAATTTGAGAGCAGCAAGCTTGAAGCTTGAAGTTCAGAAATCAGCA 1832
Db      2118 GCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCA 2177
Qy      1833 CTTCACGACAGATTCAGAGAACTTCAGCAATGCTTGGAGACTTATTCAGAGAA 1892
Db      2178 GGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAG 2237
Qy      1893 GATCTCAGTGCAGCTTAAACAAAGTCCAGCAGAACTTTGATAGAGTCTGCTTCT 1952
Db      2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2297
Qy      1953 TACCTCAAAAAGAGCTTGAAGCAAAAGAGCTGCGCCGCGCAGCAGCTCCGAGCAGCT 2012
Db      2298 GCAGAGATTTAGAGAGCAGCAGCAGCAGATTTAGAGATTCAGAGCAGATTTAGAGAGCA 2357
Qy      2013 GAGCAGGTGAGAGAGAGCAGCAGCTCAAGCTGAGAGATTTGTTTCAACAGCA 2072
Db      2358 GAGAGCAGAGTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGATTTAGAGAG 2417
Qy      2073 GCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAGCAGAGTCCCTGGA 2132
Db      2418 GCAGGAGCAGAGTTAGAGAGCAGAGCAGAGTTTGAAGAGCAGAGCAGAGATTTGA 2477
Qy      2133 GGCAGCGCAGCTGAAGCAGAAAGCAGAGAGAGAGCTGAGATTGAGAGCAAAA 2192
Db      2478 GCAGCAGCAGCAGATTTAGAGAGCAGAGCAGAGTTTGAAGAGCAGAGAGTGAAGA 2537
Qy      2193 GGAAGCGCTCAGAGAGCAGATTCAAGAAAGGAGCAAGCAATGCTGAGCATGTGCAGCA 2252
Db      2538 GCAAGAGCAGAGGTGAGAGCAGAGCAGAGCAGAGAGCAGAAATTTAGAGAGGT 2597
Qy      2253 GAGAGAGCAGCAGCCCCCGGAAACCCAGAGAGAGCAGACATGAGAGAGGAAAGCAG 2312
Db      2598 GAGAGAGCAGAGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2657
Qy      2313 TGTCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2372
Db      2658 GCAAGAGAGCAGAGAGATTTAGAGAGGTGAGAGAGCAGAGAGAGAGAGAGAGAGT 2717
Qy      2373 GCTTTTCATCCGATCAGAGCAGCAAGCTTGAAGCTGCGCAGCCAGG 2416
Db      2718 GGAAGAGCAGAGCAGCAGAGATTTAGAGAGGTGGAAGAGCAGG 2761

```

```

RESULT 14
US-09-410-399-1
; Sequence 1, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
; APPLICANT: Robertson, Etle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: to Genomic Host DNA
; FILE REFERENCE: US-03778
S

```

```

; CURRENT APPLICATION NUMBER: US/09/410,399
; CURRENT FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match      2.88; Score 144; DB 4; Length 3489;
Best Local Similarity 45.18; Pred. No. 4e-27;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

Qy      1233 GCATCAGAGGTGCTTGAAGAGCCGCTGTGAGAGATGAGCAGCAGCAGAGAAACT 1292
Db      1578 GCAGAGCCACAGAGCAGAGAGCCACAGCAGCAGAGCCACAGCAGAGAGCCACAGCA 1637
Qy      1293 GCTGTGACATTTGAAATGAAGACCGGAGAACTTGAGCCGAGCAGTGTGAGCTGGA 1352
Db      1638 GCAGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCA 1697
Qy      1353 GAAGCGCCGACAGCGCTCTTGAAGCAGCAGCCGAAAGAGCAGAGCGGTTGCTCAGCT 1412
Db      1698 GCGGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCA 1757
Qy      1413 GAGCGCGCGAGAGAGAGAGAAAGAGCGGAGAGCGCCAGAGAGCAGAGGCGCAAGCGCA 1472
Db      1758 GCGGAGCCACAGCAGCGGGAGCCACAGCAGCGGAGATGAGCAGCAGAGATGAGCAGCA 1817
Qy      1473 GCTGAGCTGAGAGAGCAGCTGAGAGAGCGGAGCTGAGAGCGGAGCAGAGAGAGAGGA 1532
Db      1818 GCAGAGATGAGAGCAGCAGATGAGCAGCAGAGATGAGAGAGCAGAGATGAGAGCA 1877
Qy      1533 GAGAGAGAGAGATCGAGAGCGCGAGCGCCGAAACGGGAATCGAAAGCAGCA 1592
Db      1878 GCAGATGAGAGCAGCAGAGATGAGCAGCAGATGAGAGCAGCAGAGATGAGAGCA 1937
Qy      1593 ACTTGAATGGGAAGGAAACCGGAGACAGGAATCTCTGAATCAGAGAAACAAGAGCAGGA 1652
Db      1938 GCAGATGAGAGCAGCAGATGAGCAGCAGCAGATGAGAGCAGCAGAGATGAGAGCA 1997
Qy      1653 GAGCAGCGTGTCTTGAAGCAGAGAGAGAGAGCTTGAAGTTTGAAGTCTTGA 1712
Db      1998 GCAGATGAGAGCAGCAGATGAGCAGCAGAGATGAGAGCAGCAGAGATGAGAGCA 2057
Qy      1713 TGACAAAAGCATCAGCTAGAGAAATTTCAAGATATCAGGTGTGATGCAATCCCA 1772
Db      2058 GCAGATGAGAGCAGCAGATGAGCAGCAGATGAGAGCAGCAGAGATGAGAGCA 2117
Qy      1773 GAGCAGAAATTTGAGAGCAGCAAGCTTGAAGAGCTTGAAGTTCAGAAATCAGCA 1832
Db      2118 GAGAGCAGAGATGAGCAGAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGCA 2177
Qy      1833 CTTCACGACAGATTCAGAGAACTTCAGCAATGCTTGAAGATTTCCAGAGAAACA 1892
Db      2178 GGAATGAGAGCAGCAGAGATGAGCAGCAGCAGCAGATGAGCAGCAGAGATGAG 2237
Qy      1893 GATCTCAGTGCAGCTTAAACAAAGTCCAGCAGAACTTTGATAGAGTCTGCTTCT 1952
Db      2238 GCAGAGCAGCAGATGAGAGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2297
Qy      1953 TACCTCAAAAAGAGCTTGAAGCAAAAGAGCTGCGCCGCGCAGCAGCTCCGAGCAGCT 2012
Db      2298 GCAGAGTTTGAAGAGCAGAGCAGAGATTTAGAGATTTAGAGAGAGAGAGAGCA 2357
Qy      2013 GAGCAGGTGAGAGAGAGCAGATCAAGCTCAGAGAGATTTGATTTTCAACAGCA 2072
Db      2358 GAGAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGAGAGAGAGAGAGAGAGAG 2417
Qy      2073 GCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAGCAGAGTCCCTGGA 2132
Db      2418 GCAGAGCAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGCAGAGATTTAG 2477

```

OY	2133	GGCAGCGGCATCTGAAAGCAGAAAGAGCAGAGAGGAAGCGCTGAGATTAGAGAAAGCAAA	2192
Db	2478	GGAGCAGGACGACGAGTTTAGAGAGCAGAGACAGAGATTAGAGAGCAGGAGGTGGAA	2537
OY	2193	GGAAAGCGCTCAGAGACGAGTTTCAGAAAAGGACAAAGCAATGCTCGAGCATGTGCAGCA	2252
Db	2538	GCAAGAGCAGGAGGTGGAAAGACAAAGACAGAGACAGGAATTAAGAGGAGGT	2597
OY	2253	GGAGAGCAGGCCACGCCCCCGGAAACCCACAGAGGAGGACAGACTGMAAGAGGAAAGCAG	2312
Db	2598	GGAGAGCGTAAGCAGCAGAGCAGAGAGACAGAGAGGAGCAGGAGTTTAGAGGAGGTGGAA	2657
OY	2313	TGTCAGGAAGAGAGGCGGAAAGAGAGACCCAAAGCCGGAATTCGAAGCAAGCAGAGTCG	2372
Db	2658	GCAGGAAGACGAGGATTAGAGAGGTGGAAAGACGAGAAAGACGAGGATTAGAGGAGGT	2717
OY	2373	GCTTTTCATCCGCACTCAGGAGCCAGCTTAAGCTGGCCACCAGG	2416
Db	2718	GGAAAGCAGAGCAGCAGAGGTTTAGAGAGAGGTGGAAAGACGAGG	2761

RESULT 15

```

US-08-770-379-20/c
; Sequence 20, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA ASSOCIATED
; NUMBER OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770.379
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3207 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-20

```

Query Match 2.8%; Score 144; DB 2; Length 32207;

Db	20419	GCAGAGACCA	OAGAGACGAG	CCACAGCA	GAGAGCC	CAACAGACAG	AGACACAGCA	20360		
OY	1293	GCCTGTCA	CACTTTG	AAGATTA	GAAGCGGG	AGAACTT	CGACCGAG	CGCAGTGTGAG	1352	
Db	20359	GCAGAGACCA	CAGCAGCGG	AGCCACAGCA	GC	CGGGAGCC	CCACAGCGGG	AGCCACAGCA	20300	
OY	1353	GAAGCGCG	CCAAAGCGCTT	TGAGAGCA	GCAGCGCA	AGCAAGAG	CGAGCGGT	TGGCTCAGCT	1412	
Db	20299	CGCGGACCC	ACAGCAGCGGG	AGCCACAGCA	GC	CGGGAGCC	ACAGCGGG	AGCCACAGCA	20240	
OY	1413	GGAGCGCG	CCAGCAGAG	AGAAAGCGGG	AGCC	CCAGAGAG	CGAGAGCG	CCAAAGCGCA	1472	
Db	20239	GC	CGGAGCC	CAACAGAGCGGG	AGCCACAGCA	GC	CGGGAGCC	ACAGCGGG	AGCCACAGCA	20180
OY	1473	GCTTGAGCT	TGAGAGAC	AGCTG	GAAGAC	CGCGAGCT	TGAGCGCG	CACGAGAGGA	1532	
Db	20179	GCAGAGAT	GAACACACAG	AGGATG	AGCAGAG	CGAGT	TGAGCAG	CGACGAGAT	20120	
OY	1533	GAGAGGA	AAAGAT	TGAGAG	GGCGG	AGCCG	AAACGGGA	CTCGAA	1592	
Db	20119	GCAGGAT	TGACACAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	20060	
OY	1593	ACTTGAAT	TGGGAA	CCGGA	CCGAG	CAGAACT	CTCGAAT	TCAGAGGA	1652	
Db	20059	GCAGGAT	TGACACAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	20000	
OY	1653	GGGCA	CGGTG	CTTGA	AGGCA	GGAGAG	AGCT	CTGAGTT	1712	
Db	19999	GCAGGAT	TGACACAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	19940	
OY	1713	TGACAAA	AACAT	CTACG	TAGAGG	AAACTT	CAGAT	TATCAG	1772	
Db	19939	GCAGGAT	TGACACAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	19880	
OY	1773	GAGGCA	GAAT	TGAGAG	CCGAA	CAAGCT	AGAGCT	TAAGAA	1832	
Db	19879	GGAGC	AGCAGAT	TAGCAG	AGGAC	AGCAT	TGACAG	CGACGAGAT	19820	
OY	1833	CTTAC	CAGCAG	AGTT	CAAGAA	TCTT	TGAGAG	CTTAT	1892	
Db	19819	GGAT	TGAGCAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	19760	
OY	1893	GAT	ACTCAG	TGACAG	TAAAC	AGTCC	AGCAAA	CAGTT	1952	
Db	19759	GCAGC	AGCAG	ACAGAT	TGAAC	AGAGCAG	AGGAG	CGAGCAG	19700	
OY	1953	TAC	CCCTCA	AAAGAG	CCCTG	GAAGCA	AGGAG	CTGCG	2012	
Db	19699	GCAGGAT	TGACACAG	ACAGGAT	TAGCAG	CGAGT	TGAGCAG	CGACGAGAT	19640	
OY	2013	GGA	CGAGT	TGAGAG	AGACCA	CGAGT	CAAGCT	CGAGAG	2072	
Db	19639	GGAGC	AGCAGAT	TAGAGAG	CGAGG	AGCAT	TGAGAG	CGACGAGAT	19580	
OY	2073	GCT	GAAGAA	CTGA	AGAGAT	CAAT	AGCAAA	CACT	2132	
Db	19579	GCAGG	AGCAGAGT	TGAGAG	CGAGCAG	AGCAT	TGAGAG	CGACGAGAT	19520	
OY	2133	GGCAG	CGCGA	CTGA	ACCA	AAAGAGAG	AGAAAG	AGCCT	2192	
Db	19519	GGAGC	AGCAG	ACAGAT	TGAGAG	CGAGG	AGCAT	TGAGAG	19460	
OY	2193	GGA	AGAC	CGCT	CAAGAG	CGAGT	TCAGAA	AGGGA	2252	
Db	19459	GCAGG	AGCAGAG	AGT	TGGAAG	AGCAAG	AGAGCAG	AGAA	19400	
OY	2253	GGAGG	AGCAG	CCAC	CGCC	CGA	AAACCC	ACAGAG	2312	
Db	19399	GGAGG	AGCAG	AAAG	ACAGAG	CGAGG	AGCAG	AGGAT	19340	
OY	2313	TGT	CAAGAA	AGAG	CGGGA	AGAG	CCAC	CGGAA	2372	
Db	19393	GCAGG	AGCAG	AGAT	TGAGAG	AGG	CTGGA	AGAG	19280	

Qy 2373 GCTTTCCATCCGATCCGAGCAGCTTAAGCTGGCCACCAGG 2416
Db 19279 GGAAGAGCAGGAGCAGAGAGCTTAAGAGAGGTGGAGAGCAGG 19236

Search completed: December 5, 2003, 00:46:34
Job time : 229.815 secs

QY 466 GATGAGAGATGATCAAGTGAATTTTCATAGCATTAGACCTTATCAAACTGAAGCTA 525
 DB 249 GATGGAAAGATGACACGACAAAGATTCTTATAGCTATGAAGCTCAAGTTAAAGTTG 308
 QY 526 CAAGATATTCAGCTCCCTCCACACTCCCTGTCATGAAACAGCAACAGTGTG--GCT 582
 DB 309 CAGGGCCAAAGACTGCTGTAGTCTCCCTCTATCAAGAAACAAACCCCTATGTTCTCT 368
 QY 583 ATTTCAGTGCACACAGCATTTGTATAGAGGGAATTGCTAGCATG-----CCACCA 633
 DB 369 CCATTAATCTGCTGCTGTTTGGGATGGGAACATGCCCAATCTGCCATTCATCAGCA 428
 QY 634 CTCACAGCTGTGCTGCTGTCGCAA-----TGGGCTCCAT 669
 DB 429 TTGCTCCAGTGCACCTATAGCAACACCTTGTCTTCTGCTACTTCAGGGAACGATTT 488
 QY 670 CCAAGTTGTGAATGCTCCACCTTATGATCTTGTCTGCTCCACAGCAGTGTCTCC 729
 DB 489 CCTCCCTTAATGATGCTGCTCCCTCAGTGCCTTCTGTTAGTACATTCCTCATTAACAAT 548
 QY 730 CTGGCTAACGGGGCTCTCCGCTCATACAGCTCTGCTGCGTTGGCGATCTGCAAGC 789
 DB 549 GGAAGCTGCAGTCTCATTCAGCCTTATTCATTCCTTATTCCTTCAACATTCGCTCAT 608
 QY 790 ACATGGCCAAAGATTCTTCTTACAGCAGATCTGGTCCAGGGTCAAAATTAACACTAG 849
 DB 609 GCATCATCTTACAGCTGATGATGGAGGATTTGGTGGCTAGTATCCAGAAAGCCAG 668
 QY 850 TTACAGAAAGCCAAATCATTCGATGTCGCGACGCGCCCTCAG----- 892
 DB 669 TCTCGATTGATTTAGGATCTAGTAGCTCACTTCTCAACGCTTCCCTCAGGGAAC 728
 QY 893 -----CAGCAGAAATGGGCTGTGCTCAGTCAATCAAGGCTGAATACAGG 936
 DB 729 TCACCTTAAGACAGGACCTCAGAGTGGGCAATTCCTCAGCCTTCAAGATTAAGATGATCGG 788
 QY 937 CAGTTATTCAGACGACGACAAACCTATAGTGGGCACTTAACAGTCCCGAGGCAAGA 996
 DB 789 CAAAAATTTAATAGTCTTAAACAAAGGCAATAGCGGATACCTCAGAGTTTCAAGCTAGA 848
 QY 997 ACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCTTCAATATGAAATCTTCT 1056
 DB 849 AATGCCCTTCTTCAAGTCAAAATCTCTCAAACTCAGCTAGCTAATTTGGACTGTGGCT 908
 QY 1057 GACATTTGATCAAGATGGAATACTCAGTGCAGAAAGATTTATCTTACGTAAGCTAAT 1116
 DB 909 GACATCGATGTGACGAGCAGTTGAAGCTGAAGAAATTTATCTGGCGATGCACTCACT 968
 QY 1117 GATGTTGCCATGTCTGGTACGACCATGCGCGCCGCTCCGCTCCAGAAATACATCCCTCT 1176
 DB 969 GACATGGCCAAAGCTGGACAGCCACTACACTGAGCTGTGCTCCGAGCTTGTCTCTCA 1028
 QY 1177 TCCCTTCAAGAGATTGCTCCGAGTGGGATGTCCTGCTCAATAGCTCTTCTTGTGAT 1236
 DB 1029 TCTTTTCAAGAG-----GGGAACCAAGTTGATTTCTGTAT 1064
 QY 1237 CAGAGCTGCTGAGAGCCGCTGCTAGAGATGAGCAGCAGCCAGAGAAAGATCGCT 1296
 DB 1065 GGAAGCTGCTTCAATATCAAGAAACACAAGAG--AAGAGCCTCAGAGAAAGCTGCA 1121
 QY 1297 GTGACATTTGAAGATTAAGAAAGGGGAGAACTTGAAGCGGAGGAGTGGAGTGAAG 1356
 DB 1122 GTTACTTTTGAAGACAAACGAAAGCCAACTATGAACGGAAGAACTGAGCTGAAG 1181
 QY 1357 CCGCCGCAAGCGCTCTTGGAGCAGCAGCCAAAGCAGAGAGCGGTGCTCAGCTGAG 1416
 DB 1182 CGAGCGCAAGTGTGATGAGCAGCAGCAGAGAGGCTGAACGCAAGCCCAAGAAAG 1241
 QY 1417 CCGCCGCAAGCAGAGAAAGCGGGAGCCCGAGAGCAGAGGCAAGCGGAGCTG 1476
 DB 1242 AAGGAAGTGGGAGCGGAAACAGAGAACTGCAAGAGCAAGAAATGGAAGAGAGCTG 1301

QY 1477 GAGCTGAGAAAGCAGCTGAGAGACAGCGGAGCTGGAGCGGACGAGAGAGAG 1536
 DB 1302 GAGTTGAGAAAGCCTTGGAGAAACAGAGAGAGCTGGAGAGACGCGGAGAGAGAG 1361
 QY 1537 AGGAAGAGATCGAGAGCGCGAGCGCAAAACGGAACCTGAAAGCAGCGCACTT 1596
 DB 1362 AGAAAGAGATGAGAAACAGAGAGGACAGAAACAGAGAGCTTGAAGACAAACCGCTT 1421
 QY 1597 GAATGGAAACGGAACCGGAGACAGAACTCCTGAATCAGAGAAACAGAGAGAGAGG 1656
 DB 1422 GAATGGAAAGACTCCCTGCGAGAGCTGCTGTAAGAAAGCAGAGAAACAAAGAGAC 1481
 QY 1657 ACCGTGCTGAGAGCAGAGAGAAAGCTCTGAGATTGAGTGAAGCTCTGAATGAC 1716
 DB 1482 ATTGTACAGCTGACCTCAGAAAGAAAGTCTCCACTGGAATCGAAGCAGTGAATGA 1541
 QY 1717 AAAAAGCATGCTAGAGAGAAACCTTCAAGATATCAGTGTGACTGGCAACCCACTTA 1776
 DB 1542 AAACATCAGCAGATCTCAGGAGACTACAAAGATGTCAAATTCAGAAAGCAAAAG 1601
 QY 1777 CAAGAAATGAGACAGAAACAGTCTAGAGAGCTAAGAAATGCTGAATCACCCTTA 1836
 DB 1602 ACTGAGCTAGAAAGTTTGTGATTAACAGTGTGACCTGGAATTAAGAAATCAACTT 1661
 QY 1837 CAGCAGAGCTGACAGAAATCTCAGCAATGCTTGAAGACTTATTCAGAGAAACAGATA 1896
 DB 1662 CAACAGAGCTTAAGAAATCAAAATTAAGCTTATCTATCTGCTCCTGAGAGCAGCTA 1721
 QY 1897 CTCAGTGAACGTTAAACAAAGTCAGCAGAAACAGTTTGATAGAGACTGCTTACC 1956
 DB 1722 TTAACGAAAGAAATTAACAAATCATGCACTCAGTAACACACTGATTCAGGGATGATT 1781
 QY 1957 CTCAAAGAGCTTGAAGAAAGAGAGCTGCGCGGAGAGCAGCTCCGGAGCAGCTGAC 2016
 DB 1782 CTTATTAAGAAAGTATGATGAAAGAAAGAAATTAATGCAAGAACTTAAGAACTTAAGAT 1841
 QY 2017 GAGGTGAGAGAGAGACAGGTCAAAGCTGACAGAGATGATGTTTCAACAAACAGCTG 2076
 DB 1842 GCTCTGAAAGAAAGAACTGATCTACGATCTCAGAAATGATTAATTAACAACTG 1901
 QY 2077 AAGAACTGAGAGATACATAGCAACAGCACTCAGAAAGCAGAGTCCCTGAGGCA 2136
 DB 1902 AAGAACTCAGAGAAAGCTATTAATACAGAGATTAACCCCTTGAACAACTTCATTAATC 1961
 QY 2137 GCGGACTGAGAGCAAGAAAGCAGAGAGAGAGCCTGAGATTAGAGAAAGCAAA 2192
 DB 1962 AAAGTGAACAAATTTGAAGAAAGAAATGAAAGAAAGATTAAGCAAAAAA 2017

RESULT 2
 US-09-907-969-72
 ; Sequence 72, Application US/09907969
 ; Publication No. US2003091580A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: King, Gordon E.
 ; APPLICANT: Algate, Paul A.
 ; APPLICANT: Piling, Steven P.
 ; APPLICANT: Rectler, Marc W.
 ; APPLICANT: Fanger, Gary Richard
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Carter, Derrick
 ; APPLICANT: Hill, Paul
 ; APPLICANT: Albone, Earl
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; FILE REFERENCE: 210121.462C8
 ; CURRENT APPLICATION NUMBER: US/09/907,969
 ; NUMBER OF SEQ ID NOS: 596
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 72

LENGTH: 2017
TYPE: DNA
ORGANISM: Homo sapiens
US-09-907-969-72

Query Match 9.5%; Score 484.8; DB 11; Length 2017;
Best Local Similarity 55.4%; Pred. No. 1.1e-128;
Matches 1128; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

226 GGGGTGATCAGCAAGTGAAGTAACTAATAGAACATGGCTCACTTCCACACTTTCGGT 285
9 GGGTGAAGCTGCAAGAAAGAGTACAGATCATATGCTCACTTCCACAGAGAAAT 68
286 GGTAGCCTGATGCTGGCCATTAATCTGAGAGAAAGGCCAAGCATGACAGAGTTC 345
69 GAGAGGCCAAATATGAGGCTATTAATCTGAGAGAAAGTAACTAATGATTAACGATT 128
346 CTTAGCCTGAAGCCGATAGCGGATTTATTAATCTGATCAAGCGAGAACTTTTTC 405
129 GATTAACCTCAACCTTCAGAGAGGTACATTAACAGTATCAAGCCGTAATTTTCC 188
406 CAATCTGGGTAACTCAAGCTCTCTTATGACAAATATGAGCGCTAGCGACATGATAC 465
189 CAGTCAGGCTCTGCGGCCCGGTTTATGAAATATGAGCCCTTATCAGATCTGAACAA 248
466 GATGAAGAGATGATCAAGTGAATTTTCCATAGCCATGAGGCTTCAACTGAAGCTA 525
249 GATGGAAGATGACCAAGCAAGGCTCTTATATGCTATGAAATCTCATCAAGTTAAAG 308
526 CAAGATATCAAGTCTCCCTCCACACTCCCTCTGATGAAACAGAACCAAGT--GCT 582
309 CAGGGCCAAAGCTGCTGATGCTCTCTCTATGATGAAACCAACCCCTATGTTCTCT 368
583 ATTTCAGAGCACACAGATTTGATATGAGAGGATGCTAGATG-----CAACA 633
369 CCACTAATCTGCTGCTGTTTGGAGTGGAAAGCATGCCAATCTGTCATTCAGCCA 428
634 CTCACAGCTGTGCTCTGTCGCA-----TGAGCTCAT 669
429 TTGCCCTCAGTTGACCTATAGCAACACCTTCTTCTGCTACTTCAGGAGCAAGTAT 488
670 CCAAGTTGTTGAATGTCCTCACCCTTATGATCTTCTGCTCCCTCAGACAGAGCTCC 729
489 CTTCCCTATATGATGCTGCTGCTCCCTATGAGCTTCTGATGATCATCTCATACCA 548
730 CTGGCTAAGGGGCTCTCCCTGATACAGCTCTGCTGCTGCTTGGCATCTGACGCC 789
549 GGAACGCAAGCTCTCATTCAGCTTATCTCATCTTATCTTCTTCAACATTTGCTCAT 608
790 ACATGSCAAAGAGTCTTCTCTCAGACATCTGTCAGAGGTCACAAATTAACATTAAG 849
609 GCATCATCTTACAGCTGATGATGAGAGATTTGGTGTGCTATGATCCAAAGGCCAG 668
850 TTACAGAGGCAATCATTCATGATGTCGACAGGCCCTCTCAG-----892
669 TCTCTGATTTAGATCTAGATCTAATCTTCTCAAGCTTCCCTCTCAGGGAAC 728
893 -----CAGCAGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 936
729 TCACCTTGAACAGGAGCTCAGAGTGGGAGTCTCAGCTTCAAGATTAAGTATCG 788
937 CAGTTATTCAGCAGCAGCAAAATATGATGAGCACTTAAACAGTCCCGCAGGAGAA 996
789 CAAAAATTTAATGCTTAAACAAAGCAGAGGATCTCTCAGAGTTTCAAGCTAAG 848
997 ACTATTTCTATGCAATCAATTTACCCAGGCTCAGCTGCTTCAATATGATGATCTTCT 1056
849 AATGCCCTTCTTCAATCTCTCAAACTCAGCTAGCTATCTTATTTGATGCTGGCT 908
1057 GACATTGATCAAGATGAGAAATCTCAGCAGAAATTTATCTTATGCTATGCACTAAT 1116
909 GACATGATGATGAGCAGCAGAGTGAAGAACTTATCTGCGATGACACTCAGCT 968

1117 GATGTGCCATGCTGTGTCAGCAGCTCGCCGCTCTCTCAGAAATATACCTCCT 1176
969 GACATGCGCAAAAGCTGACAGCCCACTACCACTGAGCTTCCCGAGTTTCTCTCA 1028
1177 TCTTCAAGAGTTCGCTCGGCAAGTGGATGTCCTGATTAAGCTCTTCTTGTGAT 1236
1029 TCTTTCAGAG-----GGGAAACCAAGTTATCTGTTAAT 1064
1237 CAGAGCTGCTGAGAGAGCCCTGTCAGAGATGAGCAGCAGCCAGAGAAATCTGCT 1296
1065 GGAACCTGCTCTTATATCAAGAAACACAAAG--AAGACCTCAGAAAGAACTGCA 1121
1297 GTACATTTGAAGATTAAGAGCGGAGAACTTCAGCAGAGGAGTGTGAGCTGAGAG 1356
1122 GTTACTTTTGAAGAACAGAAACCACTTATGAAAGAGAAATGAGCTGAGAGAG 1181
1357 GCGCCGCAAGCGCTCTTGAAGCAGAGCGCAAGAGCAGAGCGGTGCTCAGCTGAG 1416
1182 CGAGCCCAAGTGTATGAGCAGCAGCAGAGAGGCTGAAACGAAAGCCAGAAAGAG 1241
1417 CGCGCCGAGCAGAGAGAGAGAGAGCGGAGCGCCAGAGCAGAGAGCCAGCGCAG 1476
1242 AAGAAAGATGAGAGCGGAAACAGAGAACTGCAAGAGCAAGATGAGAAAGCAGCTG 1301
1477 GAGCTGAGAAACAGCTGAGAGAGAGCGGAGCTGAGCGGCGAGAGAGAGAGAGAG 1536
1302 GAGTTGAGAAACGCTTGAAGAAACAGAGAGCTGAGAGACAGCGGAGAGAGAGAG 1361
1537 AAGAAAGATGAGAGAGCGGAGCGGAGCGGAGAACTGGAAGAGCAGCAGCACTT 1596
1362 AAGAAAGATGAGAAAGCAGAGAGCAGAGCAAGAGCTTGAAGAGCAGCGGTTTA 1421
1597 GAATGGAACGGAACCGGAGAGAGAGATCTCTGATCAGAGAAACAGAGCAGAGAG 1656
1422 GAATGGAAGAGCTCGCTGCGCAGAGAGCTGCTCACTCAGAAACAGAGAGAGAG 1481
1657 ACCGTGCTCTGAGAGAGAGAGAGAGAGCTGAGAGTTGATGAGAGCTGATGAG 1716
1482 ATTTGAGCTGAGCTCCAGAAAGAAAGTCTCCAGCTGAGAGCTGAGAGAGTATG 1541
1717 AAAAAAGCATGAGTGAAGAAAGAACTTCAAGATCAAGTGTGAGCTGAGCAACCA 1776
1542 AAAAAAGCATGAGTGAAGAAAGTGAAGTGTCAAAATCAGAAAGCAACAAAG 1601
1777 CAGAAATGAGAGCAGAAAGTGTGAGAGCTTGAAGATTTGAGAAATCCCACTTA 1836
1602 ACTGAGCTAAGAGTTTGTGATTAACAGTGTGAGCTGAGAAATTAAGAAATCA 1661
1837 CAGCAGAGTTCAGAAATCTCAGAAATGCTTGAAGACTTATCCAGAGAAAGATA 1896
1662 CAACAAAGCTTAAAGAAATATCAAAATGATTTATCTGCTGCTGAGAGAGCTTA 1721
1897 CTCAAGTACAGTAAACAAAGTCCAGAGAAACAGTTTGCATAGAGACTCGTCTTACC 1956
1722 TTAACGAAAGATTTAAACATGAGCTCAGTACAGTACACACCTGATCAGGATCAG 1781
1957 CTCAAAAGAGCTTGAAGCAAGAGAGCTGCGCGCAGCAGCTCCGGAGCAGCTGAC 2016
1782 CTTCATTAAGAAAGTCAACGAAAGAGAAATTAATCCAAAGACTTAAAGAAATTA 1841
2017 GAGGTGAGAGAGAGAGCAGAGTCAAGGCTGAGAGAGTGAATTTCAACACAGCTG 2076
1842 GCTTTGAAGAAAGAACTGATCTTAAAGCTTCAAGAAATGATTTCAATTAACATCAG 1901
2077 AAGAACTGAGAGATACATAGCAAAACAGCAACTCCAGAAAGCAGAGCTCCAGAGCA 2136
1902 AAGAACTCAGAGAAAGCTATATACACAGAGTGAAGCCCTTGAACAACTTATTAAT 1961
2137 GCGCAGCTGAGCAGAAAGAGCAGAGAGAGAGAGCTGAGTGTGAGAGAGAGAA 2192
1962 AAGCTGACAAATTAAGAAATCGAAAGAAAGATTAAGCAAAAAA 2017


```

RESULT 3
US-09-827-271-72
; Sequence 72, Application US/09827271
; Publication No. US20030165504A1
; GENERAL INFORMATION:
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827,271
; CURRENT FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-827-271-72

```

Query Match	9.5%	Score 484.8;	DB 12;	Length 2017;
Best Local Similarity	55.4%;	Pred. NO. 1.1e-128;		
Match 1128; Conservative	0;	Mismatches 812;	Indels 96;	Gaps 6;

QY	226	GGCGTCGATCAAGCAAGTGAACGTATATAGAACATAGCGCTCAGATTCCACACTTTCCGT	285
Db	9	GGCTTAGAGCTGCAAGAAAGTCAAGATCATGATGGCTCAGCTTCCACAGCATGAT	68
QY	286	GGTACCTGATGATCTGGGCCATACTGTGAGAGAAAGGGCCAAAGATGACAGCATTC	345
Db	69	GGAGGGCCAAATATGTGGGCTATTACATCTGAAAGAACGTACATCAATGATTTAAACAGTTT	128
QY	346	CTTACCTGAAAGCCGATAGCCGGATTATTACTGTGTGATCAAGCAGGAACTTTTTC	405
Db	129	GATTAACCTCAAACTTTCAGAGGATTACATACAGGTGATCAAGCCCGTACTTTTTCCTA	188
QY	406	CAATCTGGATTACCTCAGCTGTCTTACGACAAATATATGGGCGCTAGCGGACATGATTAAC	465
Db	189	CAGTCAGGCTGTGCCCCGCCGTTTATAGCTGAAATATGGGCGCTTATCAGATCTGAAACAAG	248
QY	466	GATGAAAGATGAGATCAAGTGAATTTTCCATAGGCATGAACTTATCAACTGAAAGCTA	525
Db	249	GATGGAAAGATGAGACAGACAAAGTTCTTATATGCTATGAAACTCATCAAGTTAAAGTTG	308
QY	526	CAAGGATATCAAGTCCCTCCACACTTCCCTCTCATGAAACAGACAAACAAGT---	582
Db	309	CAGGGCCAAACAGCTGCTGTAGTCTCTCCCTCTATCAGAAACAAACCCCTATGTCTCT	368
QY	583	ATTTCAGTGCACACAGCATTTTGTATAGAGGAAATTGCTAGATG-----CCACCA	633
Db	369	CCACTAATCTGTGCTGTTTGGGATGGGAACATGCCCCAATCTGTCCATTGATAGCA	428
QY	634	CTCACAGCTGTGTCTCTGTGCCAA-----TGGGCTCCATT	669
Db	429	TTCCTCCAGTGCACACTATACCAACACCCTGTCTTCTGCTACTTTCAGGACACAGTAT	488
QY	670	CCAGTTTGGATGTCTTCACCCCTTAATCTTCTGTCCCTCCACAGCAAGTGCCTCC	729
Db	489	CCTCCCTTAATATGCTGTCTCCCTCAGTGCCTTCTGTATGATCAATCCATTAACAAT	548
QY	720	CTGGCTTAAGGGGCTCTCCCGTCAPACAGCTGTGCGCTTGCGCATCTGACGCC	789
Db	549	GGAACTGCCAGTCTAATTCAGCCTTATTCATCTTCTTCTTCAACATTCGCTCAT	608
QY	790	ACATGGCCAAAGATTCTTCTTCACAGCATCTGTCTCAGGGTCACAATTAAACCTAAG	849
Db	609	GCATCATCTTACAGCTGATGATGAGAGATTGGTGGTCTAGTATCCAGAAAGCCAG	668
QY	850	TTACGAAAGGCACATCAATGATGTGGCAGCGCCCTCTCAG-----	892
Db	669	TCTGTGATTGATTAGGATCTAGTAGCTCAACTTCTTCAACCTGTCCCTCAGGAAAC	728
QY	893	-----CAGCAGAAATGGGCTGTGCTCAAGTCATCAAGGCTGAATATACAG	936

Db	729	TCACCTAAGACAGGACCTCAGAGTGGCAGGTCTCTCAGCTTCAAGATTAAAGTATCGG	788
Qy	937	CAGTTATTCAACAGCCACGACMAAACTATGAGTGAACAATTAAAGGTCCCCAGGACAGA	996
Db	789	CAAAATTTAATAGCTAGTCAGACAAAGGAGATGAGCGGATATCCTCTCAGGTTTTCAAGCTAGA	848
Qy	997	ACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGGAATCTTTCT	1056
Db	849	AATGCCCTTCTTCAGTCAAACTCTCTCAAACTCAGCTAGCTACTATTATTGACCTGGCT	908
Qy	1057	GACATGTATCAAGATGGAAAATCACTAGTCAGAAAGATTATCTTAGTATGACACTAATY	1116
Db	909	GACATCGATGGTGAACGACAGTTGAAGCTGAAGATTATTTCTGGCGATCACCTCACT	968
Qy	1117	GATGTGGCATATGTGTGTGTGTCAGCCCACTGCGCCGCTCTGCTCTCCAGATACATCCCTCT	1176
Db	969	GACATGGCCAAAGCTGACACACCACTACATGACGTGGCTCCCGAGCTGTCCCTCA	1028
Qy	1177	TCCTTCAGAGAGTTCCGCTCCGCGCATGGGATGCCCTCATAGCTCTTCTTGTGGAT	1236
Db	1029	TCCTTCAGAGG-----GGAAAAGCAAGTTGATTCTGTTAAT	1064
Qy	1237	CAGAGCTGCTGAGAGACCTCCTCAGAGATGACAGACGACGAGAAATAAATCGCT	1296
Db	1065	GGAACTCTGCTTCATATCAGAAAAACACMAAAG--AAGGCTCTGAGMAAACTGCA	1121
Qy	1297	GTGACATTTGAAGATTAAGAAACCGGAGAACTTCAGCGAGGCACTGTGAGCTGAGAG	1356
Db	1122	GTTACTTTTGAAGCAAAACGGAAAGCCAACTATGAACGAGAAACATGAGCTGAGAG	1181
Qy	1357	CGCGCCAAAGCGCTCTTGGAGCAGCGCAAGCGCAAGACAGAGCGGTGGCTCAGCTGGAG	1416
Db	1182	CGACGCCAAGGTGTGATGAGACACAGACAGAGGAGGCTGAACCCAAAGCCCAAAAAGAG	1241
Qy	1417	CGCGCCGAGCAGAGAGGAAGAACCGGAGGCGCAGAGACGAGAGCCCAAGCCGACAGTG	1476
Db	1242	AAGGAAGGTGGAGCGGAAACAAGAGAACTGCAAGCAAGAAATGAGAAACAGCTG	1301
Qy	1477	GAGCTGAGAGCAGCTTGAGAAACAGCGGAGGCTGAGCGGACGACGAGAGAGAGAG	1536
Db	1302	GAGTTTGAAGAAACCGTTTGAAGAAACAGAGAGCTTGAAGACACGCGGAGAGAGAGAG	1361
Qy	1537	AGGAAGAGATCGAAGGCGCGAGGCGCAAAACGGAACTGAAAGGCAAGCGACAACTT	1596
Db	1362	AGAAAGAGATAGAAAGACGAGAGGACAAACAGGAGCTTGAAGACAAACGCCGTTTA	1421
Qy	1597	GAAATGGCAACGAAACCGGAGACAGAACTCTGAAATCAGAGAAACAGAGACAGAGGCG	1656
Db	1422	GAAATGGGAAAGACTCCGTCGCGCAGAGGTGCTCAGTCAGAGAACAGGAGAACAGAAAGAC	1481
Qy	1657	ACCGTGGCTCGAAGGCAAGGAGGAAGCTGTGAGTTTGAATTGAAGAGCTGTGAATGAC	1716
Db	1482	ATTGTCAAGGCTGAGCTCCAGAAAGAAAGTTCCACTCGAAGCTGAAAGCACTAAATGA	1541
Qy	1717	AAAAAGCATCAGCTAGAAGAAAACCTCAGAGATATCAAGTGTGCACTGGCCAGCCAGAGG	1776
Db	1542	AAACATCAGCAGATCTCAGGCAAGCTACAGATCTCCAAATCAGAAAGCAAAACAAAG	1601
Qy	1777	CAAGAAATTGAGACACGACACAGTCTAGAGAGCTTAAGAAATGCTGAATACCACTTA	1836
Db	1602	ACTGAGCTAGAAGTTTGGATTAACAGTGTGACTCGGAAATTAAGGAATTCMAAACATT	1661
Qy	1837	CAGCAGAGCTTGACGAGAACTCACAAATGTTTGAAGACTTATTCACAGAGAAACAGTA	1896
Db	1662	CAACAAAGCTTAAAGAAATCAAAATAAGTTATCTATCTGGTCCCTGAGAAACAGCTA	1721
Qy	1897	CTCAGTGCAGTTAAAAACAAGTCCAGACGAACGTTTGCATAGAGACTCGCTCTTAAC	1956
Db	1722	TTAAACGAAAGATTAAAAACATGACGCTCAGTAAACACACTGATTACAGGATCAGTTTA	1781
Qy	1957	CTCAAAAGAGCTTGAAGCAAAAGAGCTGGCCCGGACAGAGCTCCGGAGCAGCTGGAC	2016

Db 1782 CTTTCATMAAAAGTCATCAGAAAAGAAATTAATGCCAAAGACTTAAGAACAAATTAGAT 1841
Qy 2017 GAGGTGAGAGAGAGACCAAGCTCAAGCTGAGAGATGATGTTTCAACACAGCTG 2076
Db 1842 GCTCTGAAAAAAGAACTGATCTTAAGCTCTCAAAAATGATTCATTTAACTACAGCTG 1901
Qy 2077 AAGAACTGAGAGATACATAGCAAAACAGCACTCCAGAGAGAGAGCTCCGAGAGCA 2136
Db 1902 AAGAACTGAGAGAAAGCTATATATACAGCAAGTGAAGCTTGAACAACCTCAATAAAC 1961
Qy 2137 GCGCGACTGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2192
Db 1962 AAGCTGACAAATTGAAGAAATCGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2017

RESULT 4
US-10-198-053-72
Sequence 72, Application US/10198053
Publication No. US20030124140A1
GENERAL INFORMATION:
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Retter, Marc W.
APPLICANT: Fanger, Gary R.
APPLICANT: Hill, Paul
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.462C9
CURRENT APPLICATION NUMBER: US/10/198,053
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
CURRENT FILING DATE: 2002-07-17
NUMBER OF SEQ ID NOS: 624
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 72
LENGTH: 2017
TYPE: DNA
ORGANISM: Homo sapiens
US-10-198-053-72

Query Match 9.5%; Score 484.8; DB 14; Length 2017;
Best Local Similarity 55.4%; Pred. No. 1.1e-128;
Matches 118; Conservative 0; Mismatches 812; Indels 96; Gaps 6;

Qy 226 GGGGTGATCAGCAAGTGAACCTTAATAGAACCATGCTCAGTTTCCACACCTTTCGGT 285
Db 9 GGGTGAAGCTGCAAGAAAGTCAAGATCATATGCTCAGTTTCCACAGAGATGAAT 68
Qy 286 GGTAGCTGAGATCTGGGCGCATTAAGTGAAGAAAGGCGCAAGCATGACAGAGTTT 345
Db 69 GGGGGCCAAATATGTGGGCTATTACATCTGAAGAAAGTAAAGCATGATAAAGAGTT 128
Qy 346 CTTAGCTGAAGCCGATAGCGGATTTATTAATGATGATCAAGCGAGAACTTTTTC 405
Db 129 GATTAACCTCAACCTTCAGAGGTTATCAATACAGGATCAAGCCGCTTTTTCCTA 188
Qy 406 CAATCTGGGTAACTCAGCTGCTTATAGCACAATATGAGCGCTAGCGACATGAATAC 465
Db 189 CAGTCAAGTCTGCGGCGCCGCTTTAGCTGAATATGAGCTTATCAAGTCTGAACAG 248
Qy 466 GATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAGAGTTTCAAACTGAAGCTA 525
Db 249 GATGGAGATGAGACAGCAAGAGTCTCTATAGCTATGAACTCAAGTTAAAGTTG 308
Qy 526 CAAGATATCAAGTCCCTCCACACTTCCCTGATATGAAGAAAGACAGACAGT---GCT 582
Db 309 CAGGGCAACAGTCTGCTGATGCTCTCTCTATATAGAAACACCCCTATGTTCTCT 368
Qy 583 ATTTCAGTGAACACATTTGTTATAGAGAGATGCTAGATG-----CCAGCA 633
Db 369 CCACTATCTCTGCTGTTTGGGATGGGAAGCATGCCAATGTTCATATATACAGCA 428
Qy 634 CTCAAGCTGTTGCTCTCTGTCGA-----TGAGCTTCATT 669
Db 429 TTGCTCTCAAGTGAACCTATAGCAACACCTTGTCTTGTCTACTTCAAGGACCAATATT 488

Qy 670 CCAAGTTGTTGAATGTCCTCACCCTTAATGATCTTGTGCTCCCTCAGACAGAGGCTCC 729
Db 489 CTTCCCTTAATATGCTGCTCTCTCCTAGGCTTCTGTTAGTACATCTCATTTACCAAT 548
Qy 730 CTGGCTTAAGGGGCTCTCTCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTG 789
Db 549 GGAATGCAAGTCTCAATCAGCTTTATTCATCTCTTATCTTCTTCAACATTTGCTCAT 608
Qy 790 ACATGAGCAAAAGATTTCTCTTCAAGCAGATCTGCTCAGAGGTACACATTAACACTAAG 849
Db 609 GCATCATCTTAAGCCTGATGATGAGAGATTTGTGTGTCTAGTATTCAGAAAGGCCAG 668
Qy 850 TTACAGAAAGCAAAATCATTCATGATGCGCAGGCGCCCTCCAG----- 892
Db 669 TCTCTGATTTGATTTAGATCTAGTACTTCACTTCTCACTGCTTCCCTCAGGGAAC 728
Qy 893 -----CAGCAGAAATGGCTGTGCTCCTCAGTCAATCAAGCTGAATACAG 936
Db 729 TCACCTTAAGACAGGAGCCTCAGAGTGGGAGTTCTCAGGCTTCAAGATTAAGTATGG 788
Qy 937 CAGTTATTCACAGCCACAGCAAAACTATGATGACACTTAACAGTCTCCAGAGCAGAG 996
Db 789 CAAAATTTAATATGCTTGAACAAAGCATGAGCGGATACCTCTCAGGTTTCAAGCTAGA 848
Qy 997 ACTATTTCTGATCATCAAGTTTACCCAGGCTCAGGCTTCAATATGAAATCTTCT 1056
Db 849 AATGCCCTTCTCAGCAAAATCTCTCAAACTCAGCTAGCTATCTATTTGAGCTGAGCT 908
Qy 1057 GACATTTGATCAAGATGAGAAACTCACTGAGAGAAATTTATCTGATGACCTAAT 1116
Db 909 GACATGATGATGAGAGACAGTGAAGCTGAAGATTTATCTGCGCATGACCTCACT 968
Qy 1117 GATGTTCCATGCTGTGCTGACCACTGCGCCGCTCTGCTCAGAAATATCTCTCT 1176
Db 969 GACATGCGCAAAAGCTGACAGCCACTACCACTGACGTTGCTCCGAGCTTCTCTCA 1028
Qy 1177 TCCCTCAGAGAGTTGCTCCGCGAGTGGAGATGTCCTCATATAGCTCTCTCTGAT 1236
Db 1029 TCTTTCAGAG-----GGAAAGCAAGTTGATTTCTGTAT 1064
Qy 1237 CAGAGGCTGCTCAGAGAGCCGCTCTCAGAGATGACAGCAGCCAGAGAACTGCTCT 1296
Db 1065 GGAATCTGCTCTCATCTATCAAGAAACAGAGAG---AAGGCTCTCAGAGAACTGCA 1121
Qy 1297 GTGACATTTGAATGAAGAGCGGAGAACTTCAAGCAGAGCGAGTGTGAGCTGAGAG 1356
Db 1122 GTTACTTTTGAAGACAAACGAAAGCTCACTATGAACAGAGAAACATGAGCTGAGAG 1181
Qy 1357 GCGCCGCAAGCGCTCTTGAAGAGAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAG 1416
Db 1182 CGAGCGCAAGTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
Qy 1417 CCGCCGAG 1476
Db 1242 AAGAGAGATGAG 1301
Qy 1477 GAGCTGAG 1536
Db 1302 GAGTTGAG 1361
Qy 1537 AAGAGAGATGAG 1596
Db 1362 AAGAGAGATGAG 1421
Qy 1597 GAATGAG 1656
Db 1422 GAATGAG 1481
Qy 1657 ACCGTGCTCTGAG 1716
Db 1482 ATTGTCAAGCTGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1541
Qy 1717 AAAAAAGATCAGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1776

Db	1542	AAACATCAGCAGATCTCAGCGCAGATCTCAAGATGTCCAAATCCAGAAACCAACCAAAAG	1601
Qy	1777	CAAGAAATTGAGAGACCAACAAGCTCTAGAGAGCTTAAGAAITGCTGAATATCCCACTTA	1836
Db	1602	ACTGAGCTAGAAAGTTTGGATTTAAACAGTGTGACCTGGAATTTATGGAATCAACAACCTT	1661
Qy	1837	CAGCAGCACTTGCAAGGAATCTTCAGCAAAATGCTTGGAAACATTATTTCCAGAGAAACAGTA	1896
Db	1662	CAACCAAGAGCTTAAAGAAATATCCAAATTAAGCTTATCTATCTGGTCCCTGAGAAAGCAGTA	1721
Qy	1897	CTCAGTGCACGTTTAAACCAAGTCAGAGAAACAGTTTGCAATAGAGACTCGCTCTTACC	1956
Db	1722	TTTAAACGAAGAATTTAAAAACATGCAAGCTCAGTACACACCTGATTCAGGGATCAGTTTA	1781
Qy	1957	CTCAAAAGAGCCTTGGAGCAAAAGAGCTGGCCCGCAGCAGCTCCGGAGCAGCTGGAC	2018
Db	1782	CTTCTATTAATAAGTCACTCAGAAAGAGAAATTAATCCCAAGACTTTAAAGAACAAATTAAAT	1841
Qy	2017	GAGGTGAGAGAGAGACCAAGCTCAAAAGCTGCAGAGATTAATGTTTCAACAACCAAGCTG	2076
Db	1842	GCTCTTGAATAAAGAAATGCAGCTTAAGCTCTCAGAAATGGAATTCATTTAACATCAGCTG	1901
Qy	2077	AAGAACTGAGAGAGATACATATGCAAAACAGCACTCCAGAAACAGAGTCCCTGAGGCA	2133
Db	1902	AAGGAACCTCAGAGAAAGCTATATATACACAGCAGTTAGCCCTTGAACCAATTCATTAATATC	1961
Qy	2137	GCGGCACTGGAAGCAGAAAGAGCAGAGAGGAAGACCTGGAGTTAGAGAGCAAAA	2192
Db	1962	AAACCTGACAAATTTAAGGAAATTCGAAAGAAAAAATTAGACAAAAAATTTTTAAAAA	2017

RESULT 5

```

US-09-764-881-55
; Sequence 55, Application US/09764881
; Publication No. US20030125346A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PRT07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55

```

```

? LENGTH: 568
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (481)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (536)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (556)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (552)
? OTHER INFORMATION: n equals a,t,g, or c

```

Query Match	8.0%;	Score 406;	DB 11;	Length 568;
Best Local Similarity	86.2%;	Pred. No. 2.9e-106;		
Matches 480;	Conservative 0;	Mismatches 73;	Indels 4;	Gaps 3;

[illegible]

D3	75	ACCATGGCTCAGTTTCCAAACACCTTTTGGTGGCAGCCGTGATATCTGGGCCATACCTGTA	134
QY	316	GAGAAAGGCGCCAGCATGACCGACGATTTCTTAGCTGAAAGCCGATAGCGGGATTATT	375
D3	135	GAGAAAGAGCGA--CATGATCAGCAGTTCATAGTTTAAAGCCAAATATCTGGATTCATT	192
QY	376	ACGGGATCAACGCGAGAACTTTTTTCCAACTGGGGTACGTCAGCCGCTGTTAGCA	435
D3	193	ACTGGTGATCAGACTTGAATCTTTTTCATCTGGGTTAACTCTCAACTCTGTTTAGCA	253
QY	436	CAAAATATGGGCGCTAGCGGACATGATATACGATGAAAGATGATCAAGTGAATTTTCC	499
D3	253	CAATATATGGGCACTAGCTGACATGATATATATGTAAGAAATGATCAAGTGAATTTTCC	312
QY	496	ATAGCCATGAGTTATCAACTGAAGTACAGGATATCAGCTTCCCTCCACACTTCCC	555
D3	313	ATAGCTATGAACTTATCAAACTGAAGTACAGGATATCAGCTTCCCTCCACACTTCCC	372
QY	556	CCGTATGAAACAGCAACCACTGCTATTTCCAGTGCACAGGACTTTGGATATGAGAGG	615
D3	373	CTGTCTATGAACAGCAACCACTTGTATTTCTACGCCACAGCAATTGGTATGGGAGAGT	432
QY	616	ATTGTAGCATGCAACCACTCAACAGCTGTGCTCTGTGCCAATGGGCTCCATTCAGTT	679
D3	433	ATGCGCAGCAAGCAACCGCTTACACTGTGTGCTCAGTGCAGATGGGGNCATTCCAGTT	492
QY	676	GTT--GGAATGTCTCACACCTTAGTATCTTCTGTCCCTCAGACAGAGTGCTCCCTTGGC	734
D3	493	GTTGGGAATGTCTCCAAACCTTAGTATCTTCTGTCTCCACAGCA--NTGTGCCCCCTTGGC	551
QY	735	TAAAGGGGGCTCCCTCGG	751
D3	552	TAAAGGGGATCCCTTG	568

RESULT. 6

US-09-879-957-193
; Sequence 193, Application US/09879957
; Patent No. US20020034755A1

GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755a1h

KAY, Brian K.
FOWLKES, Dana M.
McCONNELL, Stephen J.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: USA
ZIP: 10036-2711

```
!      COMPUTER READABLE FORM:
!      MEDIUM TYPE: Floppy disk
!      COMPUTER: IBM PC compatible
```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 193:

SEQUENCE CHARACTERISTICS:
LENGTH: 2873 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown

MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 193:

US-09-879-957-193

Query Match 6.6%; Score 333.2; DB 9; Length 2873;

Best Local Similarity 61.0%; Pred. No. 1.1e-84;

Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

2867 CTCTCTGACCGTACCTAGTGTGCGCAGTTACGGCAGATCAGCCTTACCCAGCA 2926
541 CTTTTCACACCTTAATTAATATATATATATATATATATATATATATATATATAT 600
2927 CAGCCACTGCTCTCTCCCATCTCCGCTCTGGCCAGGTGAAAAGTGAAGGCTTAC 2986
601 TGTCCCTGG--ATCTGTATCACTTATTCATGACAGGGACAAAGTGTAGAAAACCTTA 657
2987 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAAGCAACCACTTAATTTTAACAAA 3046
658 AAGCAGAGCCCTTGTCTCTGACGTCAAGAAAGATTAACCACTTGAACCTTCAAAAC 717
3047 GTGACCTATCAACCGTCTGGAACAGCAAGACATGTGTGTGTGGAAGTTCAAGTTC 3106
718 ATGACATTTATCTGTCTTGGAGCAAGAAATGTGTGTGTGGAAGTGTGACAGAG 777
3107 AGAAGGTTGTCTCCCAAGTCTTACGTGAATCACTTCAAGGCGCCGTAGAGAAATCA 3166
778 GAAGAGATGTGTCTCCCAATCTTATGTCAAGATCATCTCTGGAGATGAATTAACGG 837
3167 CAAGATGATATCTGGCCCTACTGAAAGTCTGTAGTCAAGAAAGATGTGTCTCCCG 3226
838 AA-----GAACCAAGAGCTTTGTATGACACTGTAATTAAGAAACCTTACCTCG 885
3227 CCGCCAGGCAAGCATTTCCCGAGAGAGATTTATTCAGATATCAATATCAGAGTTCTG 3286
886 CAGCCTATTCAGT-----TGAAGAAATATATATGACATTTATCAATATCAAGTGTG 939
3287 AGCAAGAGATTTAACCTTTACAGCAAGGAGATGTGTGTGTGTACCAAGAAAGATGTG 3346
940 AACCTGAGATTTGACCTTTACAGAAAGTGAAGAAATATGTGTGACCCAGAAAGATGAG 999
3347 ACTGTGAGACGGGAGAGGTGGGAGCAAGTCCGGAATCTTCCCTTAACTATGTAGAGC 3406
1000 AGTGTGAGACGGGAGAGGTGGGAGTGAAGTGAATTTTCCATTAACCTATGTCAAC 1059
3407 TTAAGATTTACAGAGGCTCTGGAATCTGAGGAAACAGGAGATTTAGGAAAAAACTG 3466
1060 CAAGGATTCAGAGGATTTTGGAGTGTGACAGAGTGTGAGATCAATATTAACAACTG 1119
3467 AAATTTCCAGATTTATGTCTTCTACGTGTCTATCTGTCTCCCAACCACTTACCTGTG 3526
1120 AATTTCTCAGGTAATCTTACAGCATATGTGTCTTCTGTCTGAACTTAACTTAACTG 1179
3527 CTGGGAGCTGATTTCTGATTCGGGAAAGAAACCCAGATGTAGTGTGGAAGAGAACTG 3586
1180 CAGGAGATTAATTAATTTCTTAAGAAAAATTAAGAGTGTGTGCAAGGAGATTTAC 1239
3587 AAGCTCGAGGAGAAAAAGCCAGATAGGTTGTCTTCCAGCAATATTTCAACTTTCTAA 3646
1240 AAGCCAGAGAGAAAAAGCCAGAGAAAGATGTGTCTCTGCACTCATGTTAACCTTTGG 1299
3647 GCCCGGAAACAAAGCAATATCAACCCCACTGAGTCAACCAAGCCGAGTGCACGACAG 3706
1300 GTTCAAGTGTGTGAAGAGCCACACTGCTTTCATC-----CTG 1338

3707 TGGCAGGTGATCGGATGTACGATTACACCGCCAGAACGATGCGAATAGCCTTCA 3766
1339 TATGTAGGTGATTTGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 1398
3767 GCMAAGCCAGATCATCAACGTCTCAACAGAGAGACCCGAGCTGTGAAAGAGAG 3826
1399 CCAAGGACACTCATTAATGTATGTATGAACAAAGATGATCTGATGTGTGCAAGAGAG 1458
3827 TCAGTGGGCAAGTTGGGCTTCTCCATTCATTAATGTAAAGCTGACACAGACATGACC 3886
1459 TCAAGGGGCTGACTGTCTCTTCTTCCCTCAACATGATTAAGATGACGACAGCTCAGATC 1518
3887 CCAGCCAGCATTAATCA 3904
1519 CAAGTCAACATGACCA 1536

RESULT 7

US-09-764-868-125

Sequence 125 Application US/09764868

Patent No. US2002016871A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT232

CURRENT APPLICATION NUMBER: US/09/764,868

Prior application data removed - refer to PAM or file wrapper

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 125

LENGTH: 4210

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-868-125

Query Match 6.5%; Score 331.4; DB 10; Length 4210;

Best Local Similarity 61.0%; Pred. No. 4.9e-84;

Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

2867 CTCTCTGACCGTACCTAGTGTGCGCAGTTACGGCAGATCAGCCTTACCCAGCA 2926
918 CTTTTCACACCTTAATTAATATATATATATATATATATATATATATATATATAT 977
2927 CAGCCACTGCTCTCTCCCATCTCCGCTCTGGCCAGGTGAAAAGTGAAGGCTTAC 2986
978 TGTCCCTGG--ATCTGTATCACTTATTCATGACAGGAGCAAGTGTAGAAAACCTTA 1034
2987 AAGCCAGGCAAGCATTTCCCGAGAGATTTATTCAGATATCAATATCAGAGTTCTG 3046
1035 AAGCAGAGCCCTTGTCTCTGACGTCAAGAAAGATTAACCACTTGAACCTTCAAAAC 1094
3047 GTGACCTATCAACCGTCTGGAACAGCAAGACATGTGTGTGTGGAAGTTCAAGTTC 3106
1095 ATGACATTTATCTGTCTTGGAGCAAGAAATGTGTGTGTGGAAGTGTGACATGAG 1154
3107 AGAAGGTTGTCTCCCAAGTCTTACGTGAATCTCATTTCAAGGCGCCGTAGAGAAATCA 3166
1155 GAAGAGATGTGTCTCCCAATCTTATGTCAAGATCATCTCTGGAGTGAAGTAAACGG 1214
3167 CAAGATGATATCTGGCCCTACTGAAAGTCTGTAGTCAAGAGATGTGTCTCCCG 3226
1215 AA-----GAACCAAGAGCTTTGTATGACGTGTAAATTAAGAAACCTTACCTCG 1262
3227 CCGCCAGGCAAGCATTTCCCGAGAGAGATTTATTCAGATATCAATATCAGAGTTCTG 3286
1263 CAGCCTATTCAGT-----GGAAGAAATATATGTGCACTTATTCATATTCAGATGTG 1316
3287 AGCAAGAGATTTAACCTTTACAGCAAGGAGATGTATGTGTGTTCACAAAGAAATGTG 3346
1317 AACCTGAGATTTGACCTTTACAGAAAGTGAAGAAATATGTGTGACCCAGAAAGATGAG 1376

```

Oy 3347 ACTGCTGACGCGAAGCGGTGGCCGACCAAGTCTCCGAGTCTTCCCTTCTAATATGAGAGC 3406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1377 AGTGTGACAGAGAGATTTGAGATAGAGTGAATTTTTCATCAATATGCTCAAC 1436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3407 TTAAGATTCAAGAGGCTCTGGAACTGCTGGAAAAACAGGAGTTTAGAAAAAAACCTG 3466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1437 CAAGAGATCAAGAGAGTTTGGAGTGTAGCAAGTCTGAGACATCAATATAAAAACTGG 1496
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3467 AAATTGCCAGATTATGCTTCTCTACGCTGCTACGTGTGCTCCGAACTCACTCCGCTG 3526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1497 AGATTGCTCAGGTAACTTCAGCATATGTGTCTTCTGTGTGAACTTGAACCTTGTGAC 1556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3527 CTGGCAGCTGATCTGATCCGAAAAAAGAACCCAGTGTGATGTGGAGAGAGAACTGC 3586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1557 CAGACACTTAATATTAATTTCTAAAGAAAAATACAGTGTGTGTGCGAAGAGAGTTAC 1616
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3587 AAGCTGAGGAAAAAAGCGCAGATAGGGTGTTCAGCAAAATTATGTCAAACTTTCTAA 3646
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1617 AGGCGAGAGAAAAAAGCGACAGAAAGATGTGTTCTGCGCATGTAACTTTTGG 1676
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3647 GCCCGGAAACAAGCAAAATCACTCCCACTGAGCTACCAAGACCGCACTGCAGCAGAG 3706
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1677 GTCCAAAGTAGAAAGACCCACACTGCTTTTCATC-----CTG 1715
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3707 TGTGCGAGGTATCGGATGTACGATTAACCGCCAGAACGATGACGAAGCTAGCCTTCA 3766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 TATGTCAGTGTATGCTATGATGACTATGACGAAATATGATAGTACGTCTTCT 1775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3767 GCAAAGGCCAGATCATCAACGCTCTCAACAGAGAGACCCGAGCTGTGAAAGAGAGAG 3826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1776 CCAAGGACCACTATTAATGTTATGAACAAAGATGATCTGATGTTGTGGCAAGAGAGA 1835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3827 TCAGTGGCGAAGTGGGCTCTTCCATTCATTAATGTAAGTGCACACATGAGACC 3886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1836 TCAACGGGGGTGACGTGCTCTTTCTTCAAACTAGTTAAGATGACAGACTCAGATC 1895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3887 CCAGCCAGCAATG 3899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1896 CAATCAACAGTG 1908
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
US-09-918-995-31258
; Sequence 31258, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31258
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(503)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31258

```

```

Query Match 6.3%; Score 322.6; DB 11; Length 503;
Best Local Similarity 81.6%; Pred. No. 3.7e-82;
Matches 386; Conservative 0; Mismatches 84; Indels 3; Gaps 1;

```

```

Oy 1081 ACTGCAAGAAATTTATCTAGCTATGACCACTTAATGATGTGCGATCTGTGACGCA 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 AATTCAGAGAAATTTATCTGCAATGACCTCATGATGATGATGATGATGATGATGATGATG 89
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Oy 1141 CTGCGCGCCGCTCTGCTCCAGAAATACATCCCTTCTCTTCAAGAGAGTTGCTCCGCGC 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 CTGCGACCTGTCTGCTCTCCAGAAATACATTCACCTTCTTTTAAAGAGTTTGTGATCTGCGC 149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1201 AGTGGATGTCCTCATAGCTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGCTG 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 AGTGTATATCTGTCAATAGTCAACATCTGTATACAGAGGCTTACCAAGAGAACAGTT 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1261 TCAGAGATGACGACGAC--CAGAGAGAAATGCTGCTGTGACATTTGAATTAAGAG 1317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 TTAAGATGAAACAACAATTTAGAAAAGAAATTAACCTGTAAAGTTGAAGATTAAGAG 269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1318 CGGAGAACTTCGAGCGAGGAGATGTGAGCTGTGAGAGAGCGCCGCAAGCGCTTGTGAG 1377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 CGGAGAACTTTGAACCTGTGCACTGCACTGAACTGAGAAACGAGCAAGCTCTCTGGA 329
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1378 CAGCAGCGCAAAAGACAGAGCCGTTGCTCACTGAGCGCGCGACAGAGAGAGAAA 1437
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 CAGCAGCGCAAGAGAGAGAGCGCTGCTGAGCGCGCGAGCGAGAGAGAGAGAGAG 389
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1438 GAGCGGAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 GAGCGTACGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 449
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 1498 AAGCAGCGGAGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1550
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AAGCAGCGGAGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 9
US-09-879-957-39
; Sequence 39, Application US/09879957
; Patent No. US2002003475A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US2002003475A1h
; KAY, Brian K.
; FOWLER, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:

```

QY	255	GTGGATGAAAGCCAGATGGAGAGCCAGATGGCTTGGAGAGAGCTGAAAGGGAACAGC	2616
Db	1	GTGGATGAAAGCCAAATGAGAAACCCGCTGGCTTGGAGAGAGATTAAAGGAAGACAA	60
QY	2617	GGATGGTTCCTCTGAAACTATGACAGAAAGATTTCCAAATAATAGCTTCCATCTCCAGCC	2676
Db	61	GGGTGGTTCCTCTGAAACTATGACAGAAAGATTTCCAAATAATAGCTTCCGCTCCAGTG	120
QY	2677	AAACCAATGACCAATCTGACATCTGCCCCCTGCCCCCAACTGGCTCTGCGTAGAGCCCTT	2736
Db	121	AAACCAATGACTAATTCACATCTGCCCCCTGCCCCCAACTGGCTCTGCGTAGAGACCCCT	180

Qy 2737 GCTCCTTGCCAGTACCTCTTCTGAGCCCTCCACACCCCAACACTGGGAGACTTC 2796
 Db 181 GCCCTTTGGCAGTAACTCTTCCAGAGCCCTCCACAGACCTTAATTAATCTGGGCGGACTTC 240
 Qy 2797 AGTTCACGT 2806
 Db 241 AGTTCACGT 250

RESULT 11
 US-09-864-761-26948
 ; Sequence 26948, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmicsa-X-1
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 26948
 ; LENGTH: 286
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AF000311.1
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.87
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
 ; OTHER INFORMATION: NT HIT: AF114488.1, EVALUATE 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
 ; OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUATE 1.00e-112
 US-09-864-761-26948

Query Match 3.9%; Score 198.8; DB 9; Length 286;
 Best Local Similarity 87.2%; Pred. NO. 1.7e-46;
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 2557 GTGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAGACG 2616
 Db 1 GTGATGAAAGCCAACTGGAGAACCCGGCTGCTTGGAGAGAGATTAAGAAACACA 60
 Qy 2617 GGATGTTCCCTGCAAACTATGCAAAAGATTCCAGAAAATGAGTTCCACTCCAGCC 2676
 Db 61 GGGTGGTTCCTGCAAACTATGCAAGAAAATCCAGAAAATGAGTTCCCTCCAGTG 120
 Qy 2677 AAACAGTACCATCTGACATGCGCCCTGCCCCCAAGCTGCTGCGAGACCCCT 2736
 Db 121 AAACAGTACCATCTGACATGCGCCCTGCCCCCAAGCTGCTGCGAGACCCCT 180
 Qy 2737 GCTCCTTGCCAGTACCTCTTCTGAGCCCTCCACACCCCAACACTGGGAGACTTC 2796
 Db 181 GCCCTTTGGCAGTAACTCTTCCAGAGCCCTCCACAGACCTTAATTAATCTGGGCGGACTTC 240
 Qy 2797 AGTTCACGT 2806
 Db 241 AGTTCACGT 250

RESULT 12
 US-09-864-761-30453
 ; Sequence 30453, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmicsa-X-1
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30


```

PRIORITY APPLICATION NUMBER: 2001-01-30
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/668,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30453
LENGTH: 297
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HB100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: AF114487.1, EVALUATE 0.00e+00
OTHER INFORMATION: EST HUMAN HIT: AA73263.1, EVALUATE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
US-09-864-761-30453

```

Db 181 AACTGAGAGCAATGTGAAAAAGCGCGGAGCTTGAACCGCAGAGAGAGAGAGAGA 240
Qy 1538 GGAAGAGATGAGAGCGCGCGGAGCGCGGAAAAAGCGGA 1574
Db 241 GGAAGAGAAATTGAGAGCGGAGGAGTGTACGAGCGCGAGA 277

RESULT 14
US-09-864-761-17644
Sequence 17644, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17644
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUATE 3.00e-62
OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 2.80e-01
OTHER INFORMATION: NT HIT: AF114487.1, EVALUATE 1.00e-122
US-09-864-761-17644

Query Match 3.9%; Score 197.4; DB 9; Length 263;
Best Local Similarity 84.4%; Pred. No. 4e-46;
Matches 222; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1298 TGACATTGAAGATAGAGAGCGGAGAACTTCGAGCGGACAGTGTGAGCTGGAGAGC 1357
Db 1 TAACTTGAAGATAGAGAGCGGAGAACTTGAAGCTGGCACTGGAAGAAAC 60
Qy 1358 GCGCCAGCGCTCTTGAAGAGAGCGGCAAGAGAGAGAGCGGCTTCACTGAGAGC 1417
Db 61 GAGGCAAGCTCTCTGGAACAGCAGCGCAAGAGAGAGAGAGCGGCTTGGCCAGCTGGAGC 120
Qy 1418 GCGCCAGCGGAGAGAGAGAGAGCGGAGCGGCAAGAGAGAGAGAGCGGAGAGCTGG 1477
Db 121 GGGGAGAGAGAGAGAGAGAGAGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 1478 AGCTGAG 1537
Db 181 AACTGAG 240
Qy 1538 GGAAGAGATGAGAGAGCGCGGAG 1560
Db 241 GGAAGAGAAATTGAGAGCGGAGAG 263

RESULT 15
US-09-864-761-10314
Sequence 10314, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30

```

1  PRIOR APPLICATION NUMBER: PCT/US01/00661
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: PCT/US01/00670
4  PRIOR FILING DATE: 2001-01-30
5  PRIOR APPLICATION NUMBER: US 60/234,687
6  PRIOR FILING DATE: 2000-09-21
7  PRIOR APPLICATION NUMBER: US 09/608,408
8  PRIOR FILING DATE: 2000-06-30
9  PRIOR APPLICATION NUMBER: US 09/774,203
10 PRIOR FILING DATE: 2001-01-29
11 NUMBER OF SEQ ID NOS: 49117
12 SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
13 SEQ ID NO 10314
14 LENGTH: 480
15 TYPE: DNA
16 ORGANISM: Homo sapiens
17 FEATURE:
18 OTHER INFORMATION: MAP TO AP000311.1
19 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
20 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
21 OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
22 OTHER INFORMATION: EXPRESSED IN FETAL LAYER, SIGNAL = 32
23 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
24 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
25 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
26 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
27 OTHER INFORMATION: EXPRESSED IN ADULT LAYER, SIGNAL = 0.98
28
29 US-09-864-761-10314

```

Query Match	3.8%	Score 192	DB 9;	Length 480;
Best Local Similarity	81.6%	Pred. No. 2.2e-44;		
Matches 222; Conservative	0;	Mismatches 50;	Indels 0;	Gaps 0;

OY	2553	TGAGATCAACATCCGACGAGAGATATGTATGTGTGTATGAAGCCAGATCGAGAGACC	2582
Db	209	TTAGTCTACTTAAAGCTGTGACTTTTTCACAGGTGTGATGAAGCCAAATCGAGAAC	268
OY	2583	AGATGCGTTTGAGAGAGAGCTGAAGAGAGACGGAGATGTTTCCTGCAAACTATGACGA	2642
Db	269	CGGCTGCGTTTGAGAGAGAAATTTAAAAAGAAAGACAGGGTGTTCTCTGCAAACTATGACGA	328
OY	2643	AAAGATTCCAGAAAAATGAGTTCCCACTCCAGCCAAACAGGACCGATCTGACATCTGCG	2702
Db	329	GAAATTCACGAAATAATGAGGTTCCCGCTCCAGTGAACCAAGTGAATTCAACATCTGCG	388
OY	2703	CCCTGCCCCCAACTGCTGTGCGTGAGAACCCCTGCTCTTTGCCAGTGAACCTCTTTCGA	2762
Db	389	CCCTGCCCCCAACTGCTGTGCGTGAGAACCCCGCCTTTGGCAGTAACTCTTTCGA	448
OY	2763	GGCCTCCCAACCCCCCAACCAACTGGGCGAGACT	2794
Db	449	GGCCTCCCAACCCCCCTAATAACTGGGGCGAGCT	480

Search completed: December 5, 2003, 06:15:39
Job time : 1089.45 secs

DR P-PSDB; AAY57444.
 XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection
 XX
 PS Claim 6; Page 40-42; 9pp; English.
 XX
 CC The present sequence encodes mouse Ese1. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (1) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (1) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (1)-specific antibodies (Ab); sequences antisense to the (1)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (1) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

Query Match 100.0%; Score 3642; DB 21; Length 3723;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTGATTTCCACACCTTTCGGTGTAGCTGTGATGTCGTGGCCATTAATCTGAG 60
 DB 1 ATGGCTGATTTCCACACCTTTCGGTGTAGCTGTGATGTCGTGGCCATTAATCTGAG 60
 QY 61 GAAAGGGCCAGCATGACGACGAGTCTTACCTGAAAGCCGATAGCGGATTTATTA 120
 DB 61 GAAAGGGCCAGCATGACGACGAGTCTTACCTGAAAGCCGATAGCGGATTTATTA 120
 QY 121 GGTATCAAGGAGGAACTTTTTCATATCTGGGTTAATCTCAACCTGCTTACGACA 180
 DB 121 GGTATCAAGGAGGAACTTTTTCATATCTGGGTTAATCTCAACCTGCTTACGACA 180
 QY 181 ATATGGGCGCTAGCGGATGATTAACGATGGAAGATGATCAATGGAATTTTCCATA 240
 DB 181 ATATGGGCGCTAGCGGATGATTAACGATGGAAGATGATCAATGGAATTTTCCATA 240
 QY 241 GCCATGAAGCTTATCAAACTGAAGCTAGAGATATCAAGCTCCCTCCACATTTCCCT 300
 DB 241 GCCATGAAGCTTATCAAACTGAAGCTAGAGATATCAAGCTCCCTCCACATTTCCCT 300
 QY 301 GTCTGAAACAGCAACCACTGCTATTTCCAGTGCACACGATTTGGATAGAGAGATT 360
 DB 301 GTCTGAAACAGCAACCACTGCTATTTCCAGTGCACACGATTTGGATAGAGAGATT 360
 QY 361 GTCTGATGCAACCACTGCTATTTCCAGTGCACACGATTTGGATAGAGAGATT 420
 DB 361 GTCTGATGCAACCACTGCTATTTCCAGTGCACACGATTTGGATAGAGAGATT 420
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGCTCTCCAGAGACAGTGTCTCCCTGGCTAAC 480
 DB 421 GGAATGTCTCCACCTTATGATCTTCTGCTCTCCAGAGAGAGTGTCTCCCTGGCTAAC 480
 QY 481 GGGGCTCTCTCCCTGATACAGCTCTGCTGCTTGGGATCTCTGACGCCACATGGCCA 540
 DB 481 GGGGCTCTCTCCCTGATACAGCTCTGCTGCTTGGGATCTCTGACGCCACATGGCCA 540
 QY 541 AAGAGTTCTTCTTCCAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTTACAGAG 600

DB 541 AAGAGTTCTTCTTCCAGAGATCTGGTCCAGGGTCAACAATTAAACATTAAGTTACAGAG 600
 QY 601 GCACAAATCATTTGATGTGCGAGCGCCCTTCACAGCAGAAATGGCTGTGCTCACTCA 660
 DB 601 GCACAAATCATTTGATGTGCGAGCGCCCTTCACAGCAGAAATGGCTGTGCTCACTCA 660
 QY 661 TCAAGGCTGAAATCAGGAGTTTATTCACAGCCACGACAAATATGATGATGACACTTA 720
 DB 661 TCAAGGCTGAAATCAGGAGTTTATTCACAGCCACGACAAATATGATGATGACACTTA 720
 QY 721 ACAGGTCGCCAGGACAAATATTCATGCAATCAAGTTTACCCAGGCTCACTGAGCT 780
 DB 721 ACAGGTCGCCAGGACAAATATTCATGCAATCAAGTTTACCCAGGCTCACTGAGCT 780
 QY 781 TCAATATGAAATCTTTTGTGATGATCAAGATGAGAAAATCACTGCAAGAAATTTATC 840
 DB 781 TCAATATGAAATCTTTTGTGATGATCAAGATGAGAAAATCACTGCAAGAAATTTATC 840
 QY 841 CTAGCTATGCACTTAATTTGATGTTGCAATGTCTGTCAGCCACGCGCCGCTGCTGCT 900
 DB 841 CTAGCTATGCACTTAATTTGATGTTGCAATGTCTGTCAGCCACGCGCCGCTGCTGCT 900
 QY 901 CCAGAAATACATCCCTCTCTCTTCCAGAAAGTTGCTCCGCACTGAGATGTCCGTGATA 960
 DB 901 CCAGAAATACATCCCTCTCTCTTCCAGAAAGTTGCTCCGCACTGAGATGTCCGTGATA 960
 QY 961 AGCTCTTCTTGTGATTCAGAGGCTGCTGAGAGCCGCTGTGAGATGACAGACAG 1020
 DB 961 AGCTCTTCTTGTGATTCAGAGGCTGCTGAGAGCCGCTGTGAGATGACAGACAG 1020
 QY 1021 CCAGAGAAAGAACTGCTGTGATCAATTTGAAATGAAAGCGGAGAACTTCCAGCAGAGC 1080
 DB 1021 CCAGAGAAAGAACTGCTGTGATCAATTTGAAATGAAAGCGGAGAACTTCCAGCAGAGC 1080
 QY 1081 AGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTTGAAGCAGCAGCCGCAAGACAGAG 1140
 DB 1081 AGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTTGAAGCAGCAGCCGCAAGACAGAG 1140
 QY 1141 CGTGTGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAAAGCGGAGCGCCAGAGAGCAG 1200
 DB 1141 CGTGTGCTCAGCTGAGAGCGCGCCGAGCAGAGAGAAAGCGGAGCGCCAGAGAGCAG 1200
 QY 1201 GAGGCGCAAGCGGAGCTGAGAGCTGAGAGAGAGTGAAGAGCGGAGAGCTGAGAGCGG 1260
 DB 1201 GAGGCGCAAGCGGAGCTGAGAGCTGAGAGAGAGTGAAGAGCGGAGAGCTGAGAGCGG 1260
 QY 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAACGGGAACTG 1320
 DB 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCGCAAAACGGGAACTG 1320
 QY 1321 GAAAGGCAAGCAACTTGAATGGGAAACGAAACCGAGACAGAACTCTGATATCAGAGG 1380
 DB 1321 GAAAGGCAAGCAACTTGAATGGGAAACGAAACCGAGAGACAGAACTCTGATATCAGAGG 1380
 QY 1381 AACAAG 1440
 DB 1381 AACAAG 1440
 QY 1441 TTAGAAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAAATCTTCAAGATATCAGGTG 1500
 DB 1441 TTAGAAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAAATCTTCAAGATATCAGGTG 1500
 QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAGAAAGTCTAGAGAGCTTAAGAAAT 1560
 DB 1501 CGACTGGCAACCCAGAGGCAAGAAATTTGAGAGCAGAAAGTCTAGAGAGCTTAAGAAAT 1560
 QY 1561 GCTGAATATCCCACTTAACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 DB 1561 GCTGAATATCCCACTTAACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 ATTCCAGAGAAACGATATCTGATGACCAAGTTAAACCAAGTCCAGCAGAAACAGTTTCAT 1680

Db	1621	ATTCCAGAGAAACAAGATATCTCATGTATCCAGTTTAAACAAGTCCAGACAGAAACAGTTTGCAT	1680
OY	1681	AGAGACTCGCTTTCTTACCTCTTAAAGAGCCTTTGGAGCAAGAGACTGCGCGACAG	1740
Db	1681	AGAGACTCGCTTTCATCCCTCAAAAAGACCTTTGGAGCAAAAGAGACTGCGCGACAG	1740
OY	1741	CTCCGGGAGCAGCTGGACGAGGTGGAGAGAGACCAAGTCTAAAGCTGACAGACATTTGAT	1800
Db	1741	CTCCGGGAGCAGCTGGACGAGGTGGAGAGAGACCAAGCTGACAGACATTTGAT	1800
OY	1801	GTTTTCAACAACACAGCTGAAGAACTGAGAGAGATACATPATCAAAACACCAACTCCAGAG	1860
Db	1801	GTTTTCAACAACACAGCTGAAGAACTGAGAGAGATACATPATCAAAACCAACTCCAGAG	1860
OY	1861	CAGAGGTCCTTGAGAGCAGCGGACCTGAAGCAGAAAGACAGAGAGAGAGAGCCTGGAG	1920
Db	1861	CAGAGGTCCTTGAGAGCAGCGGACCTGAAGCAGAAAGACAGAGAGAGAGAGCCTGGAG	1920
OY	1921	TTAGGAAGCAAAAAGAGAGCGCTCTAGAGACGATTCAGGAAAGGGAACACATATGCTG	1980
Db	1921	TTAGGAAGCAAAAAGAGAGCGCTCTAGAGACGATTCAGGAAAGGGAACACATATGCTG	1980
OY	1981	GAGCATGTGACAGCAGAGAGAGCAGCGCCCGGAAACCCCAACAGAGAGACAGACTG	2040
Db	1981	GAGCATGTGACAGCAGAGAGAGCAGCGCCCGGAAACCCCAACAGAGAGAGACAGACTG	2040
OY	2041	AAGAGGGAAGACAGTGTCCAGAGAAAGAGGCGGAAAGAGAGAGCCAGCCGAAATGCA	2100
Db	2041	AAGAGGGAAGACAGTGTCCAGAGAAAGAGGCGGAAAGAGAGAGCCAGCCGAAATGCA	2100
OY	2101	GACCAAGCAGAGTCGGCTTTTCCATCCGATCAGAGCCAGCTAAGCTGGCCACCCAGCA	2160
Db	2101	GACCAAGCAGAGTCGGCTTTTCCATCCGATCAGAGCCAGCTAAGCTGGCCACCCAGCA	2160
OY	2161	CCCTGGTATCAACAAGAAAGAGGCGGCTTACATTTCTGACAGAGAGATGTAAAGTG	2220
Db	2161	CCCTGGTATCAACAAGAAAGAGGCGGCTTACATTTCTGACAGAGAGATGTAAAGTG	2220
OY	2221	GTATATTATCCGAGCGCTGTACCCCTTTGATTCAGAACTCAAGATGATCAATCCAG	2280
Db	2221	GTATATTATCCGAGCGCTGTACCCCTTTGATTCAGAACTCAAGATGATCAATCCAG	2280
OY	2281	CCAGAGATATATGTCATGTGTGATGAAAGCCAGACTGGAGAGCCAGATGCTTGAGGA	2340
Db	2281	CCAGAGATATATGTCATGTGTGATGAAAGCCAGACTGGAGAGCCAGATGCTTGAGGA	2340
OY	2341	GAGCTGAAAGGGAAGACGGGATGTGTTCCCTGCAAACTATGCAAAAGATTCAGAAAT	2400
Db	2341	GAGCTGAAAGGGAAGACGGGATGTGTTCCCTGCAAACTATGCAAAAGATTCAGAAAT	2400
OY	2401	GAGGTTCCCACTCCAGCCAAACACAGATGACCATCTGATCGCCCTGCCCCCAACTG	2460
Db	2401	GAGGTTCCCACTCCAGCCAAACACAGATGACCATCTGATCGCCCTGCCCCCAACTG	2460
OY	2461	GCTCTGGGTGAGACCCCTGTCTCTTTGCAAGTGACTTTCTTGAGCCCTTCACAACCCC	2520
Db	2461	GCTCTGGGTGAGACCCCTGTCTCTTTGCAAGTGACTTTCTTGAGCCCTTCACAACCCC	2520
OY	2521	AAACAATGGGAGAGATTCAAGTTCCAGTGGGCCACAGACTCAAAAGAGAACCCAGAAACG	2580
Db	2521	AAACAATGGGAGAGATTCAAGTTCCAGTGGGCCACAGACTCAAAAGAGAACCCAGAAACG	2580
OY	2581	GACCACTGGGATACCTGGGCGGCTCAGCCTTCTGACCCGTATCCTATGTCCTGGCCAGTTA	2640
Db	2581	GACCACTGGGATACCTGGGCGGCTCAGCCTTCTGACCCGTATCCTATGTCCTGGCCAGTTA	2640
OY	2641	CGGCAAGATCAGCTTTTATCCCGACCAACAGCACTGGCTCTCTCCCATCTCCGTCCTG	2700
Db	2641	CGGCAAGATCAGCTTTTATCCCGACCAACAGCACTGGCTCTCTCCCATCTCCGTCCTG	2700
OY	2701	GCGCAGGGTGAAGAAAGTGGAAAGGCTACAAAGCGCAAGCCCTGTATCTCTGGAGAGCCAA	2760
Db	2701	GCGCAGGGTGAAGAAAGTGGAAAGGCTACAAAGCGCAAGCCCTGTATCTCTGGAGAGCCAA	2760

Oy	276	AAAGACAAACCACTTAATTTTAAACAAAAGAGAGCTATCAACCGTCTCGAAACAGCAAGAC	2820
Dp	2761	AAAGACAAACCACTTAATTTTAAACAAAAGAGAGCTATCAACCGTCTCGAAACAGCAAGAC	2820
Oy	2821	ATGTGTGTGTTGGAGAGTTCAAGCTCAGAGGGTGTGTTCCCAAGTCTTAGCGTAAA	2880
Dp	2821	ATGTGTGTGTTGGAGAGTTCAAGGTCAGAGGGTGTGTTCCCAAGTCTTAGCGTAAA	2880
Oy	2881	CTCATTTTCAGAGGCCCGTAAAGAAATCCACAAGCATCGATCTGGCCCTACTGAAGTCTCT	2940
Dp	2881	CTCATTTTCAGAGGCCCGTAAAGAAATCCACAAGCATCGATCTGGCCCTACTGAAGTCTCT	2940
Oy	2941	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCAGCCAGCCATTTCCCGAGAGAAAGATT	3000
Dp	2941	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCAGCCAGCCATTTCCCGAGAGAAAGATT	3000
Oy	3001	ATTGCCATGTACACATACGAGAGTTCTGAGCAAGAGATTAACTTTACGACAGAGGGAT	3060
Dp	3001	ATTGCCATGTACACATACGAGAGTTCTGAGCAAGAGATTAACTTTACGACAGAGGGAT	3060
Oy	3061	GTCATTTGTGGTTACCAAGAAAGATGTGACTGTGTGACCGGAAACGTTGGGCGACAAATCC	3120
Dp	3061	GTCATTTGTGGTTACCAAGAAAGATGTGACTGTGTGACCGGAAACGTTGGGCGACAAATCC	3120
Oy	3121	GGAGTCTTCCCTCTAACTATGAGAGGCTTAAAGATTCAAGAGGCTGTGAACTGCTGGG	3180
Dp	3121	GGAGTCTTCCCTCTCTAACTATGAGAGGCTTAAAGATTCAAGAGGCTGTGAACTGCTGGG	3180
Oy	3181	AAAACAGGAGATTGAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTTAGCGTCT	3240
Dp	3181	AAAACAGGAGATTGAGAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTTAGCGTCT	3240
Oy	3241	ACTGTGTCCGAAACAACCTCACCTTGGCTCTTGGGCGAGCTGATTTCTGATCCGGAAAAAGAAC	3300
Dp	3241	ACTGTGTCCGAAACAACCTCACCTTGGCTCTTGGGCGAGCTGATTTCTGATCCGGAAAAAGAAC	3300
Oy	3301	CCAGGTGATGTGTGTGAGAGAGAACTGCAAGCTCGAGGAGAAAAAGCGCCAGATTTAGGGTGG	3360
Dp	3301	CCAGGTGATGTGTGTGAGAGAGAACTGCAAGCTCGAGGAGAAAAAGCGCCAGATTTAGGGTGG	3360
Oy	3361	TTTTCCAGCAAAATTATGTCAAACTTTCTTAAAGCCCGGAAACAAGCAAAATCACCCCAACTGAG	3420
Dp	3361	TTTTCCAGCAAAATTATGTCAAACTTTCTTAAAGCCCGGAAACAAGCAAAATCACCCCAACTGAG	3420
Oy	3421	CTAACCCAAAGCCGAGTGCAGGCCAGCAGTGTGCTCGAGTATCGGAGTGTACGATTACACC	3480
Dp	3421	CTAACCCAAAGCCGAGTGCAGGCCAGCAGTGTGCTCGAGTATCGGAGTGTACGATTACACC	3480
Oy	3481	GCCCGAAGACGATGACGAACCTAGCCCTTACGAAAGGCGAGATTCATCAACGCTCTCAACAG	3540
Dp	3481	GCCCGAAGACGATGACGAACCTAGCCCTTACGAAAGGCGAGATTCATCAACGCTCTCAACAG	3540
Oy	3541	GAGGACCCCGGACTGTGTGAAAGAGAAAGTCAAGTGTGGGCAAGTTGGGCTTTCCATCCAAAT	3600
Dp	3541	GAGGACCCCGGACTGTGTGAAAGAGAAAGTCAAGTGTGGGCAAGTTGGGCTTTCCATCCAAAT	3600
Oy	3601	TATGTAAAGCTGACCAACAGACATGTGACCCCAAGCCAGCAATGA 3642	
Dp	3601	TATGTAAAGCTGACCAACAGACATGTGACCCCAAGCCAGCAATGA 3642	
RESULT 2			
AAZ39025			
ID	AAZ39025	standard; cDNA; 5144 BP.	
XX	AAZ39025;		
XX	AC		
XX	DT	28-FEB-2000 (first entry)	
XX	DE	Mouse Esei1 coding sequence.	
XX	XX	Mouse; murine; Esei1; Esei2; endocytosis; vesicular trafficking;	

KW regulation: actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX Mus sp.
 PN M0955728-A2.
 XX 04-NOV-1999.
 PD 27-APR-1999; 99WO-CA00375.
 XX 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX (HSCR-) HSC RES & DEV LP.
 PA Egan SE, Wang W, Sengar A;
 PI MPI, 2000-052802/04.
 DR P-PDB; AAY57449.
 XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 XX Claim 6; Page 59-62; 9pp; English.
 XX The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Eps15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese1 coding sequence.
 CC
 XX
 SQ Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;
 Query Match 99.9%; Score 3639.4; DB 21; Length 5144;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db	241	GCATGAAGCTTATCAAACTGAAGCTACAAAGAAATCAGCTCCCTCCACACTTCCCT	300
Qy	301	GTCTAATAACAGCAACCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGGAT	360
Db	301	GTCTAATAACAGCAACCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGGAT	360
Qy	361	GCTAGCAATGACCACTCAGAGCTGTGCTCTGTCGCAATGGAGCTCATTCAGATTGT	420
Db	361	GCTAGCAATGACCACTCAGAGCTGTGCTCTGTCGCAATGGAGCTCATTCAGATTGT	420
Qy	421	GGAATGCTCCACCCCTTATGATCTTCTGCTCCCTCCAGACAGAGTCTCCCTGGCTAAC	480
Db	421	GGAATGCTCCACCCCTTATGATCTTCTGCTCCCTCCAGACAGAGTCTCCCTGGCTAAC	480
Qy	481	GGGGCTCTCCCTGCTATACAGACCTCTGCTGCTTGGCGATCTCGACCAATGGCCA	540
Db	481	GGGGCTCTCCCTGCTATACAGACCTCTGCTGCTTGGCGATCTCGACCAATGGCCA	540
Qy	541	AAGAGTCTTCTCCAGACAGTCTGCTCCAGGGTCAAAATTAACATTAAGTTACAGAG	600
Db	541	AAGAGTCTTCTCCAGACAGTCTGCTCCAGGGTCAAAATTAACATTAAGTTACAGAG	600
Qy	601	GCACATCATTCGATGTCGCGAGCGCCCTCCAGACAGAGATGGGCTGCTCACTCA	660
Db	601	GCACATCATTCGATGTCGCGAGCGCCCTCCAGACAGAGATGGGCTGCTCACTCA	660
Qy	661	TCAGGCTGAATATACAGAGCTTATTTCAACAGCCACGACAAATTAAGTGGACACTTA	720
Db	661	TCAGGCTGAATATACAGAGCTTATTTCAACAGCCACGACAAATTAAGTGGACACTTA	720
Qy	721	ACAGGCTCCAGAGAAAGTATCTCATGCAATCAAGTTAACCCAGGCTCAGTGGCT	780
Db	721	ACAGGCTCCAGAGAAAGTATCTCATGCAATCAAGTTAACCCAGGCTCAGTGGCT	780
Qy	781	TCATATGGAATCTTCTGACATTCATGATCAAGATGAAGAACTCAGTCGAGAAATTAATC	840
Db	781	TCATATGGAATCTTCTGACATTCATGATCAAGATGAAGAACTCAGTCGAGAAATTAATC	840
Qy	841	CTAGCTATGACCTTAATGATGTTGTCATGTCGTCAGCCACTGCGGCTCTGCT	900
Db	841	CTAGCTATGACCTTAATGATGTTGTCATGTCGTCAGCCACTGCGGCTCTGCT	900
Qy	901	CCAGATATACCTCCCTCTCCAGAAAGTGGCTCCGCGAGTGGAGATGCCGTCA	960
Db	901	CCAGATATACCTCCCTCTCCAGAAAGTGGCTCCGCGAGTGGAGATGCCGTCA	960
Qy	961	AGCTCTTCTTCTGATGATCAAGAGCTGCTGAGAGCGCTGTCAGAGATGAGCAGAG	1020
Db	961	AGCTCTTCTTCTGATGATCAAGAGCTGCTGAGAGCGCTGTCAGAGATGAGCAGAG	1020
Qy	1021	CCAGAGAAAGAACTGCTGTAATTAAGATTAAGAAAGCCGGGAACCTTGACCGAGGC	1080
Db	1021	CCAGAGAAAGAACTGCTGTAATTAAGATTAAGAAAGCCGGGAACCTTGACCGAGGC	1080
Qy	1081	AGTGTGAGCTGGAAGCGCGCAAGCGCTCTTGAAGCAGCGCAAGAGAGAG	1140
Db	1081	AGTGTGAGCTGGAAGCGCGCAAGCGCTCTTGAAGCAGCGCAAGAGAGAG	1140
Qy	1141	CGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAAAGCGCGAGCGCAGAGAG	1200
Db	1141	CGTTGGCTCAGCTGAGCGCGCGAGAGAGAGAAAGCGCGAGCGCAGAGAG	1200
Qy	1201	GAGGCAAGCGGAGCTGAGAGCTGGAAGAGCAGCTGGAAGAGCGGAGAGCGG	1260
Db	1201	GAGGCAAGCGGAGCTGAGAGCTGGAAGAGCAGCTGGAAGAGCGGAGAGCGG	1260
Qy	1261	CAGGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGGCAAAAGCGGAACTG	1320
Db	1261	CAGGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGGCAAAAGCGGAACTG	1320
Qy	1321	GAAAGGAGAGCAACTTGAATGGAAAGGAAAGCGGAGACAGAACTCTGAATCAGAGG	1380
Db	1321	GAAAGGAGAGCAACTTGAATGGAAAGGAAAGCGGAGACAGAACTCTGAATCAGAGG	1380

Db 1321 GAAAGGCAACGACCTTGAATGGAACGGAACCGAGACAGGAACCTCTGAATCAGAG 1380
Qy 1381 AACAGAGCAGAGAGGCAACCGTGTCTGAAGGCAAGAGAGAGCTTGAAGTTGAG 1440
Db 1381 AACAGAGCAGAGAGGCAACCGTGTCTGAAGGCAAGAGAGAGCTTGAAGTTGAG 1440
Qy 1441 TTGAGAGCTCTGAATGACAAAAGCATCAGTGAAGAAAATTGAGATATCAGGTGT 1500
Db 1441 TTGAGAGCTCTGAATGACAAAAGCATCAGTGAAGAAAATTGAGATATCAGGTGT 1500
Qy 1501 CGACTGGAACCCAGAGGCAAGAAATTGAGAGCAAGCAAGTCTAGAGAGCTAAGAAAT 1560
Db 1501 CGACTGGAACCCAGAGGCAAGAAATTGAGAGCAAGCAAGTCTAGAGAGCTAAGAAAT 1560
Qy 1561 GGTGAATACCCACTTACAGAGCAGTGAAGAAATCTCAGAAATGCTTGAAGAACTT 1620
Db 1561 GGTGAATACCCACTTACAGAGCAGTGAAGAAATCTCAGAAATGCTTGAAGAACTT 1620
Qy 1621 ATTCCAGAGAAACAGATCTCAGTGAAGCAAGTTAAACAGTCCAGAGAAAGTTGAT 1680
Db 1621 ATTCCAGAGAAACAGATCTCAGTGAAGCAAGTTAAACAGTCCAGAGAAAGTTGAT 1680
Qy 1681 AAGAGCTGCTTCTTACCTCAAAGAGCTTGAAGCAAGAGCTGAGCCGAGCAG 1740
Db 1681 AAGAGCTGCTTCTTACCTCAAAGAGCTTGAAGCAAGAGCTGAGCCGAGCAG 1740
Qy 1741 CTCGGAGAGCAGTGAAGAGGTGAAGAGAGCAAGTCAAGTCAAGTCTGAGAGATGAT 1800
Db 1741 CTCGGAGAGCAGTGAAGAGGTGAAGAGAGCAAGTCAAGTCAAGTCTGAGAGATGAT 1800
Qy 1801 GTTTTCAACACAGCTGAAGAACTGAGAGATCATAGCAACAGCACTCCAGAG 1860
Db 1801 GTTTTCAACACAGCTGAAGAACTGAGAGATCATAGCAACAGCACTCCAGAG 1860
Qy 1861 CAGAGTCCCTGAGAGCAGCGGCACTGAAGCAAGAAAGCAGAGAGAGAGAGCTGAG 1920
Db 1861 CAGAGTCCCTGAGAGCAGCGGCACTGAAGCAAGAAAGCAGAGAGAGAGAGCTGAG 1920
Qy 1921 TTGAGAGCAAGAAAGAAAGCTCAGAGAGAGTCAAGAAAGGAGCAAGCAATGAGCTG 1980
Db 1921 TTGAGAGCAAGAAAGAAAGCTCAGAGAGAGTCAAGAAAGGAGCAAGCAATGAGCTG 1980
Qy 1981 GAGCATGTGACAGAGAGAGAGCAGCAGCCCGGAAACCCCAAGAGAGAGCAGACTG 2040
Db 1981 GAGCATGTGACAGAGAGAGAGCAGCAGCCCGGAAACCCCAAGAGAGAGCAGACTG 2040
Qy 2041 AAGAGGAAAGACAGTGTCAAGAAAGAGAGGCGGAAGAGAGAGCCAGCCGAAATGCA 2100
Db 2041 AAGAGGAAAGACAGTGTCAAGAAAGAGAGGCGGAAGAGAGAGCCAGCCGAAATGCA 2100
Qy 2101 GACAGAGAGTGTGCTTTTCCATCCGATCAGAGAGCAGTGAAGTGTGAGCCAGAGCA 2160
Db 2101 GACAGAGAGTGTGCTTTTCCATCCGATCAGAGAGCAGTGAAGTGTGAGCCAGAGCA 2160
Qy 2161 CCTGTCTACCAAGAGAAAGGCGCTTACCATTTGCAAGAGAGTGAAGAAAGT 2220
Db 2161 CCTGTCTACCAAGAGAAAGGCGCTTACCATTTGCAAGAGAGTGAAGAAAGT 2220
Qy 2221 GTATATTCAGAGCGCTGTACCTCTTTGAATCCAGAAAGTCAAGATGATCAACATCCAG 2280
Db 2221 GTATATTCAGAGCGCTGTACCTCTTTGAATCCAGAAAGTCAAGATGATCAACATCCAG 2280
Qy 2281 CCAGAGATATATGTCTATGTGTGAAGAAAGCAGAGCTGAGAGCAGAGATGAGTGAAG 2340
Db 2281 CCAGAGATATATGTCTATGTGTGAAGAAAGCAGAGCTGAGAGCAGAGATGAGTGAAG 2340
Qy 2341 GAGCTGAAGAGAGAGAGAGAGAGTGTCTCTGCAAACTATGCAAGAAAGTTCAAGAAAT 2400
Db 2341 GAGCTGAAGAGAGAGAGAGAGTGTCTCTGCAAACTATGCAAGAAAGTTCAAGAAAT 2400
Qy 2401 GAGGTTCCACTTCAGCCAAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2460
Db 2401 GAGGTTCCACTTCAGCCAAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2460

Qy 2461 GCTTGGGTGAAGACCCCTGCTCTTTGCAAGTGAAGTCTTCTGAGCCCTCCAGACCC 2520
Db 2461 GCTTGGGTGAAGACCCCTGCTCTTTGCAAGTGAAGTCTTCTGAGCCCTCCAGACCC 2520
Qy 2521 AACAACTGGGAGACTTCAAGTTCCAGTGGCCAGAGCACTCAAGAGAGAGCAAGAG 2580
Db 2521 AACAACTGGGAGACTTCAAGTTCCAGTGGCCAGAGCACTCAAGAGAGAGCAAGAG 2580
Qy 2581 GACAACTGGGATACGTTGAGGCGGCTCAGGCTTCTGACCGTACCTAGTGTGGGCAAGTT 2640
Db 2581 GACAACTGGGATACGTTGAGGCGGCTCAGGCTTCTGACCGTACCTAGTGTGGGCAAGTT 2640
Qy 2641 CGGAGAGATCAGCCCTTTACCCAGCCAGCAGCAGTGTCTCTCCCATCTCCGCTCTG 2700
Db 2641 CGGAGAGATCAGCCCTTTACCCAGCCAGCAGCAGTGTCTCTCCCATCTCCGCTCTG 2700
Qy 2701 GGCAGAGGTGAAGAGGAGGCTACAGGCGCAAGCCCTGTATCCCTGAGAGCCAA 2760
Db 2701 GGCAGAGGTGAAGAGGAGGCTACAGGCGCAAGCCCTGTATCCCTGAGAGCCAA 2760
Qy 2761 AAAGCAACCACTTAATTTTAAACAAAGTGAAGTCAAGTCTGAGAGCAAGAG 2820
Db 2761 AAAGCAACCACTTAATTTTAAACAAAGTGAAGTCAAGTCTGAGAGCAAGAG 2820
Qy 2821 ATGTGTGTGTGAGAGAGTCAAGTCAAGAGAGGTTGGTTCCCAAGTCTTACGTGAA 2880
Db 2821 ATGTGTGTGTGAGAGAGTCAAGTCAAGAGAGGTTGGTTCCCAAGTCTTACGTGAA 2880
Qy 2881 CTCATTTAGAGGCGCGTGAAGAAATCCAGAGCATCATCTGAGCCCTAAGAAAGTCT 2940
Db 2881 CTCATTTAGAGGCGCGTGAAGAAATCCAGAGCATCATCTGAGCCCTAAGAAAGTCT 2940
Qy 2941 GCTAGTCTTAAAGAGAGTGTCTTCCCGGCGCAAGCCAGCATTCGCGGAGAGAGTT 3000
Db 2941 GCTAGTCTTAAAGAGAGTGTCTTCCCGGCGCAAGCCAGCATTCGCGGAGAGAGTT 3000
Qy 3001 ATTGCAATGTACATCAAGAGTGTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 3060
Db 3001 ATTGCAATGTACATCAAGAGTGTGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAG 3060
Qy 3061 GTGATGTGTGTTACCAAGAAAGTGTGAGTGTGAGAGGGAACGTTGGGAGCAAGTCC 3120
Db 3061 GTGATGTGTGTTACCAAGAAAGTGTGAGTGTGAGAGGGAACGTTGGGAGCAAGTCC 3120
Qy 3121 GAGGTCTTCCCTTCAATATGTAGAGGCTTAAAGATTCAGAGGCTCTGAACTGCTGG 3180
Db 3121 GAGGTCTTCCCTTCAATATGTAGAGGCTTAAAGATTCAGAGGCTCTGAACTGCTGG 3180
Qy 3181 AAAAGAGGAGTTTGAAGAAAAAACCTGAATTTGCCAGGTTATTTCTTCTAGCTGCT 3240
Db 3181 AAAAGAGGAGTTTGAAGAAAAAACCTGAATTTGCCAGGTTATTTCTTCTAGCTGCT 3240
Qy 3241 ACTGTCTCCGAACCACTCAGCCCTGCTCTGAGGAGCTGATTCGATCCGGAAGAAAG 3300
Db 3241 ACTGTCTCCGAACCACTCAGCCCTGCTCTGAGGAGCTGATTCGATCCGGAAGAAAG 3300
Qy 3301 CCAGGTGATGTGTGAAGAGAACTGCAAGCTGAGGGAAGAAAGCCAGATGAGTGG 3360
Db 3301 CCAGGTGATGTGTGAAGAGAACTGCAAGCTGAGGGAAGAAAGCCAGATGAGTGG 3360
Qy 3361 TTTCCAGCAAAATATATGTCAAACTTTAAGCCCGGAACAGCAAAATATCACTGAG 3420
Db 3361 TTTCCAGCAAAATATATGTCAAACTTTAAGCCCGGAACAGCAAAATATCACTGAG 3420
Qy 3421 CTACCAAGACCGAGTGAAGAGAGAGTGTGAGAGGAGTGAAGTGAAGTGAAGTGAAG 3480
Db 3421 CTACCAAGACCGAGTGAAGAGAGAGTGTGAGAGGAGTGAAGTGAAGTGAAGTGAAG 3480
Qy 3481 GGCAGAGAGTGAAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 3540
Db 3481 GGCAGAGAGTGAAGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG 3540

Query Match 99.9%; Score 3639.4; DB 21; Length 5738;

[illegible]

Db 1280 CAGAGAAAGAACTGCTGTGACATTGAAATAGAGCGGAGAACTTGACGAGGC 1339
QY 1081 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGAGAGAGCGCAAGAGCGAG 1140
Db 1340 AGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGAGAGAGCGCAAGAGCGAG 1399
QY 1141 CGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1200
Db 1400 CGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1459
QY 1201 GAGGCCAAGCGGAGAGCTGAGAGAGCTGAGAGAGCAAGCGGAGAGCTGAGCGG 1260
Db 1460 GAGGCCAAGCGGAGAGCTGAGAGAGCAAGCTGAGAGAGCAAGCGGAGAGCTGAGCGG 1519
QY 1261 CAGCGAGAGAGAGAGAGAGATCGAGAGGCGCGCAAAACGGGAACTG 1320
Db 1520 CAGCGAGAGAGAGAGAGAGATCGAGAGGCGCGCAAAACGGGAACTG 1579
QY 1321 GAAAGGCAAGCACTTGAATGGGAAACGGAAACGGAACTCTGTAATCAGAG 1380
Db 1580 GAAAGGCAAGCACTTGAATGGGAAACGGAAACGGAACTCTGTAATCAGAG 1639
QY 1381 AACCAAGAGCAGAGAGGCGACCGTGTCTGAGAGCAAGAGAACTCTGAGTTGAG 1440
Db 1640 AACCAAGAGCAGAGAGGCGACCGTGTCTGAGAGCAAGAGAACTCTGAGTTGAG 1699
QY 1441 TTAGAAGCTCTGAATGACAAAGATCAGCTAGAGGAAATCTTGAAGATCAGGTGT 1500
Db 1700 TTAGAAGCTCTGAATGACAAAGATCAGCTAGAGGAAATCTTGAAGATCAGGTGT 1759
QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAGCTAGAGAGCTAAGAT 1560
Db 1760 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAACAGCTAGAGAGCTAAGAT 1819
QY 1561 GCTGAATACCACTTACAGCAGCAGTTGCGAGAAATCTCAGCAATGCTTGGAGACT 1620
Db 1820 GCTGAATACCACTTACAGCAGCAGTTGCGAGAAATCTCAGCAATGCTTGGAGACT 1879
QY 1621 ATTCCAGAGAAAGATCAGTGAAGTGAAGCAAGTTAAACAGTCCAGAGAAACGTTTGCAT 1680
Db 1880 ATTCCAGAGAAAGATCAGTGAAGTGAAGCAAGTTAAACAGTCCAGAGAAACGTTTGCAT 1939
QY 1681 AAGAGCTCGCTTCTTAAACCTTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCAGCAG 1740
Db 1940 AAGAGCTCGCTTCTTAAACCTTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCAGCAG 1999
QY 1741 CTCGGGAGCAGCTGAGCAGAGTGAAGAGAGACCAAGTCAAAAGCTGACAGAGATTGAT 1800
Db 2000 CTCGGGAGCAGCTGAGCAGAGTGAAGAGAGACCAAGTCAAAAGCTGACAGAGATTGAT 2059
QY 1801 GTTTTCAACAACAGCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAA 1860
Db 2060 GTTTTCAACAACAGCTGAAGAACTGAGAGAGATCATAGCAAAACAGCACTCCAGAA 2119
QY 1861 CAGAGTCCCTGAGGCGCAGCTGAAGCAGAAAGAGCAGAGAGAGAGAGCTGAGAG 1920
Db 2120 CAGAGTCCCTGAGGCGCAGCTGAAGCAGAAAGAGCAGAGAGAGAGAGCTGAGAG 2179
QY 1921 TTAGAGACCAAAAGAGAGCGCTCAGAGACGAGTTCAAGAAAGGAGCAAGCAATGCTG 1980
Db 2180 TTAGAGACCAAAAGAGAGCGCTCAGAGACGAGTTCAAGAAAGGAGCAAGCAATGCTG 2239
QY 1981 GAGCATGTGACAG 2040
Db 2240 GAGCATGTGACAG 2299
QY 2041 AAGAGGAAAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2300 AAGAGGAAAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359
QY 2101 GACAGCAGAGTGTGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGAGCAGCAGGCA 2160
Db 2360 GACAGCAGAGTGTGCTTTTCCATCCGATCAGAGAGCAGCTAAGCTGAGCAGCAGGCA 2419

QY 2161 CCTGGTCTACAG 2220
Db 2420 CCTGGTCTACAG 2479
QY 2221 GTATATTAACGAGCGGTGTACCTTTGAATCCAGAAAGTCAAGATGATCAATCCAG 2280
Db 2480 GTATATTAACGAGCGGTGTACCTTTGAATCCAGAAAGTCAAGATGATCAATCCAG 2539
QY 2281 CCAAGAGATATGTCTATGTGTGATGAAGCCAGACTGAGAGCAGAGATGAGCTTGGAG 2340
Db 2540 CCAAGAGATATGTCTATGTGTGATGAAGCCAGACTGAGAGCAGAGATGAGCTTGGAG 2599
QY 2341 GAGCTGAAG 2400
Db 2600 GAGCTGAAG 2659
QY 2401 GAGGTTCCCACTCCAGCCAAACAGTGAACGATCTGACATCTGAGCCCTCCCAAACTG 2460
Db 2660 GAGGTTCCCACTCCAGCCAAACAGTGAACGATCTGACATCTGAGCCCTCCCAAACTG 2719
QY 2461 GCTCTGCTGAGAGCCCTGCTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
Db 2720 GCTCTGCTGAGAGCCCTGCTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2779
QY 2521 AACAACTGGGAGAGCTTCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
Db 2780 AACAACTGGGAGAGCTTCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2839
QY 2581 GACAACTGGGATAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2640
Db 2840 GACAACTGGGATAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2899
QY 2641 CGCAGAGATCAGCTTTTACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2700
Db 2900 CGCAGAGATCAGCTTTTACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2959
QY 2701 GGCAGAGGAGAAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
Db 2960 GGCAGAGGAGAAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3019
QY 2761 AAGAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAATCAAGCTCTGAGAGAGAG 2820
Db 3020 AAGAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAATCAAGCTCTGAGAGAGAG 3079
QY 2821 ATGTGAGTGTGAGAGAGTCAAGGTCAAGAGGAGTGTGAGTGTGAGTGTGAGTGT 2880
Db 3080 ATGTGAGTGTGAGAGAGTCAAGGTCAAGAGGAGTGTGAGTGTGAGTGTGAGTGT 3139
QY 2881 CTCATTTCAAGGCGCGTGAAGAAATCCACAGCATCGATCTGAGGCTTCTGAAAGTCT 2940
Db 3140 CTCATTTCAAGGCGCGTGAAGAAATCCACAGCATCGATCTGAGGCTTCTGAAAGTCT 3199
QY 2941 GCTAGCTTAAAGAGAGTGTGCTCCCGGCGGCAAGCAGCAGCTTCCGAGAGAGT 3000
Db 3200 GCTAGCTTAAAGAGAGTGTGCTCCCGGCGGCAAGCAGCAGCTTCCGAGAGAGT 3259
QY 3001 ATTGCATGTACATCAGAGAGTCTGAGCAGAGAGATTAACTTTCAAGAGAGAGAT 3060
Db 3260 ATTGCATGTACATCAGAGAGTCTGAGCAGAGAGATTAACTTTCAAGAGAGAGAT 3319
QY 3061 GTGATTTGTGTATCCAAAGAAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 3120
Db 3320 GTGATTTGTGTATCCAAAGAAAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAG 3379
QY 3121 GAGGCTTCCCTTCTTAACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGAGAGCTG 3180
Db 3380 GAGGCTTCCCTTCTTAACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGAGAGCTG 3439
QY 3181 AAAACAGAGAGTTAG 3240
Db 3440 AAAACAGAGAGTTAG 3499

Oy 3241 ACTGTCGCCGAACAACCTCAGCCCTGCTCTGAGCAGTGAATTCGATCCGGAAAAAGAAC 3300
 Db 3500 ACTGTCGCCGAACAACCTCAGCCCTGCTCTGAGCAGTGAATTCGATCCGGAAAAAGAAC 3559
 Oy 3301 CCAGGTGATGATGGGGAAGAGAACTGCAAGCTCGAGGGAAAAAGCCGAGATAGGCTGG 3360
 Db 3560 CCAGGTGATGATGGGGAAGAGAACTGCAAGCTCGAGGGAAAAAGCCGAGATAGGCTGG 3619
 Oy 3361 TTTCCAGCAAAATTAATGTAACCTTTAAGCCCGGAAACAGCAAAATATACCCCAATGAG 3420
 Db 3620 TTTCCAGCAAAATTAATGTAACCTTTAAGCCCGGAAACAGCAAAATATACCCCAATGAG 3679
 Oy 3421 CTACCAAGACCGCAGTGCAGCAGCAGTGTGCTCAGTGTATGCGGATGATACATTACAC 3480
 Db 3680 CTACCAAGACCGCAGTGCAGCAGCAGTGTGCTCAGTGTATGCGGATGATACATTACAC 3739
 Oy 3481 GCCCAGAACGATGACGAACCTAGCCTTCAGCAAAAGCCAGATATCAACGCTCTCAACAG 3540
 Db 3740 GCCCAGAACGATGACGAACCTAGCCTTCAGCAAAAGCCAGATATCAACGCTCTCAACAG 3799
 Oy 3541 GAGAACCCGAGTGTGGAAGAGAAAGTCAATGAGGCAAGTTGGCTTTCCCATCCAT 3600
 Db 3800 GAGAACCCGAGTGTGGAAGAGAAAGTCAATGAGGCAAGTTGGCTTTCCCATCCAT 3859
 Oy 3601 TATGTAAGCTGACACAGACATGAGACCCAGCCAGCAATG 3641
 Db 3860 TATGTAAGCTGACACAGACATGAGACCCAGCCAGCAATG 3900

RESULT 4

AA239008
 ID AA239008 standard; cDNA; 5082 BP.

AA239008;

28-FEB-2000 (first entry)

Mouse Eae1 full length cDNA sequence.

Mouse; murine; Eae1; Eae2; endocytosis; vesicular trafficking;
 regulation; actin cytoskeleton; detection; cancer; infection;
 EH-domain and SH3-domain regulator of endocytosis; anticancer;
 antiproliferative; antiviral; ss.

OS Mus sp.

PN W09955728-A2.

PD 04-NOV-1999.

PF 27-APR-1999; 99MO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

PA (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A;

DR WPI; 2000-052802/04.

DR P-PSDB; AAY57444.

PT New nucleic acid encoding Eae1 and 2 proteins, involved in regulation

PT of endocytosis, used e.g. for treating cancer or preventing viral

PT infection

PS Claim 6; Page 38-40; 99pp; English..

CC The present sequence encodes mouse Eae1. The present invention

CC specifically describes mammalian Eae1 and 2 proteins (I) and their splice

CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)

CC are involved in regulation of clathrin-mediated endocytosis (as a complex

CC with Eap1s protein), vesicular trafficking and actin cytoskeleton.

CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants): (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Eae genes or
 CC antagonists of an Eae binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Eae1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Eae mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Eae-Esp1s complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.

Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;

Query Match 99.7%; Score 3630; DB 21; Length 5082;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;

Matches 3641; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Oy 1 ATGAGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATGCTGGCCATTACTGTGAG 60
 Db 259 ATGAGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATGCTGGCCATTACTGTGAG 318
 Oy 61 GAAAGGGCCAGACATGACACAGCACTTCTTACCTGAAAGCCGATAGGGGATTTATTA 120
 Db 319 GAAAGGGCCAGACATGACACAGCACTTCTTACCTGAAAGCCGATAGGGGATTTATTA 378
 Oy 121 GGTATCAGAGGAGAACTTTTTCATCTGGGTTTACCTGAGCTGCTTGTAGCAAA 180
 Db 379 GGTATCAGAGGAGAACTTTTTCATCTGGGTTTACCTGAGCTGCTTGTAGCAAA 438
 Oy 181 ATATGGCGCTAGGCGACATGAATTAACATGAGAGATGATGATCAAGTGAATTTTCATA 240
 Db 439 ATATGGCGCTAGGCGACATGAATTAACATGAGAGATGATGATCAAGTGAATTTTCATA 498
 Oy 241 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATATAGTCTCCCTCCACACTTCCCT 300
 Db 499 GCCATGAAGCTTATCAAACTGAAGCTACAGAGATATAGTCTCCCTCCACACTTCCCT 558
 Oy 301 GTCATGAAGACGCAACAGAGTGGCTATTTCAGTGCACAGCATTTGGTATAGAGGGATT 360
 Db 559 GTCATGAAGACGCAACAGAGTGGCTATTTCAGTGCACAGCATTTGGTATAGAGGGATT 618
 Oy 361 GCTAGCATGCCACCACTACAGCTGTGCTCTGTGCAATGGGCTCATTCAGATTGT 420
 Db 619 GCTAGCATGCCACCACTACAGCTGTGCTCTGTGCAATGGGCTCATTCAGATTGT 678
 Oy 421 GGAATGTCCTCAACCTTAATATCTTCTGCTCCCTCAGACAGAGTGCCTCCCTGGCTAAC 480
 Db 679 GGAATGTCCTCAACCTTAATATCTTCTGCTCCCTCAGACAGAGTGCCTCCCTGGCTAAC 738
 Oy 481 GGGGCTCCTCCGTCATACAGCCTCTGCTGCTGGGCTTGGGCTCTGAGCCACATGGCCA 540
 Db 739 GGGGCTCCTCCGTCATACAGCCTCTGCTGCTGGGCTTGGGCTCTGAGCCACATGGCCA 798
 Oy 541 AAGAGTCTTCTTCCAGCAGATCTGTCTCAGGGTCAACATTTAACTAATTTACAGAG 600
 Db 799 AAGAGTCTTCTTCCAGCAGATCTGTCTCAGGGTCAACATTTAACTAATTTACAGAG 858
 Oy 601 GCAACATCATTCATATGTCGCGAGGCCCTTCACAGACAGAAATGGCTGCTCAGTCA 660
 Db 859 GCAACATCATTCATATGTCGCGAGGCCCTTCACAGACAGAAATGGCTGCTCAGTCA 918
 Oy 661 TCAAGGCTGAATATCAGCAGTATTATCAACAGCCAGCAAAATATGATGAGTGAACATTA 720
 Db 919 TCAAGGCTGAATATCAGCAGTATTATCAACAGCCAGCAAAATATGATGAGTGAACATTA 978
 Oy 721 ACAGGTCGCCAGGAGAACTATTTCTCATGCAATCAAGTTTACCAGGCTCAAGCTGGCT 780

Db 979 AAGAGTCCCGAGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGTGGCT 1038
QY 781 TCAATATGGAATCTTCTGACATTTGATCAAGATGAAATCTGCTGCAAGAAATTTATC 840
Db 1039 TCAATATGGAATCTTCTGACATTTGATCAAGATGAAATCTGCTGCAAGAAATTTATC 1098
QY 841 CTAGCTATGCACTTAAATGATGTTGTCATGCTGTCAGCACTGCGCCGCTGCTGCT 900
Db 1099 CTAGCTATGCACTTAAATGATGTTGTCATGCTGTCAGCACTGCGCCGCTGCTGCT 1158
QY 901 CCAAGATACATCCCTCTCTCTCTTCAAGAGATTCGCTCCGAGAGTGGATGTCCTCAT 960
Db 1159 CCAAGATACATCCCTCTCTCTCTTCAAGAGATTCGCTCCGAGAGTGGATGTCCTCAT 1218
QY 961 AGCTCTTCTCTGTCATGATGAGAGTGGCTGTCAGAGACCCGCTGTCAGAGATGTCAGAG 1020
Db 1219 AGCTCTTCTCTGTCATGATGAGAGTGGCTGTCAGAGACCCGCTGTCAGAGATGTCAGAG 1278
QY 1021 CCAGAGAGAACTGCTGCTGTCATTTGAAATGATGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1279 CCAGAGAGAACTGCTGCTGTCATTTGAAATGATGAGAGAGAGAGAGAGAGAGAGAG 1338
QY 1081 AGTGTGAGCTGAG 1140
Db 1339 AGTGTGAGCTGAG 1398
QY 1141 CGTTGGCTCAGCTGAG 1200
Db 1399 CGTTGGCTCAGCTGAG 1458
QY 1201 GAGGCAAG 1260
Db 1459 GAGGCAAG 1518
QY 1261 CAGCGAG 1320
Db 1519 CAGCGAG 1578
QY 1321 GAAAGGCAAG 1380
Db 1579 GAAAGGCAAG 1638
QY 1381 AACCAAG 1440
Db 1639 AACCAAG 1698
QY 1441 TTAGAGAGCTCTGATGATCAAAAAGATCAGCTAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1699 TTAGAGAGCTCTGATGATCAAAAAGATCAGCTAGAGAGAGAGAGAGAGAGAGAG 1758
QY 1501 CGAGCTGAG 1560
Db 1759 CGAGCTGAG 1818
QY 1561 GCTGAATACACCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1819 GCTGAATACACCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
QY 1621 ATTCCAG 1680
Db 1879 ATTCCAG 1938
QY 1681 AAGAGATCGCTTTTACCTTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1939 AAGAGATCGCTTTTACCTTCAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1998
QY 1741 CTCGGAG 1800
Db 1999 CTCGGAG 2058
QY 1801 GTTTTCAACAAACAGCTGAG 1860
Db 2059 GTTTTCAACAAACAGCTGAG 2118

QY 1861 CAGAGGTCCCTGAG 1920
Db 2119 CAGAGGTCCCTGAG 2178
QY 1921 TTAG 1980
Db 2179 TTAG 2238
QY 1981 GAGCATGTGACAG 2040
Db 2239 GAGCATGTGACAG 2298
QY 2041 AAG 2100
Db 2299 AAG 2358
QY 2101 GACAG 2160
Db 2359 GACAG 2418
QY 2161 CCTGGTCTACAG 2220
Db 2419 CCTGGTCTACAG 2478
QY 2221 GATATATTCAG 2280
Db 2479 GATATATTCAG 2538
QY 2281 CCAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
Db 2539 CCAGAGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2598
QY 2341 GAGCTGAAG 2400
Db 2599 GAGCTGAAG 2658
QY 2401 GAGGTTCCCACTCCAG 2460
Db 2659 GAGGTTCCCACTCCAG 2718
QY 2461 GCTTGGCTGAG 2520
Db 2719 GCTTGGCTGAG 2778
QY 2521 AACCACTGGAG 2580
Db 2779 AACCACTGGAG 2838
QY 2581 GACCACTGGAG 2640
Db 2839 GACCACTGGAG 2897
QY 2641 CGGAG 2700
Db 2899 CGGAG 2957
QY 2701 GGCAG 2760
Db 2959 GGCAG 3017
QY 2761 AAG 2820
Db 3018 AAG 3077
QY 2821 ATGT 2880
Db 3078 ATGT 3137
QY 2881 CTCATTTCAAG 2940
Db 3138 CTCATTTCAAG 3197


```

QY 2241 GCTAGCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCATTCCCGAGAAAGATT 3000
Db 3198 GCTAGCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCATTCCCGAGAAAGATT 3257
QY 3001 ATTGCGATGACATACGAGAGTTCTGAGCAAGAGATTAACTTTCACCAAGGGAT 3060
Db 3258 ATTGCGATGACATACGAGAGTTCTGAGCAAGAGATTAACTTTCACCAAGGGAT 3317
QY 3061 GTGATGTTGTTACCAAGAAAGATGTCAGTGGGAGGGAAGGTTGGGAGCAAGTCC 3120
Db 3318 GTGATGTTGTTACCAAGAAAGATGTCAGTGGGAGGGAAGGTTGGGAGCAAGTCC 3377
QY 3121 GAGAGTCTTCCCTTCTAATATGATGAGGCTTAAAGTTGAGAGGGCTCTGAACTGCTGGG 3180
Db 3378 GAGAGTCTTCCCTTCTAATATGATGAGGCTTAAAGTTGAGAGGGCTCTGAACTGCTGGG 3437
QY 3181 AAAACAGGAGTTTNGAAAAAACCTGAAATTGGCCAGTTTATTGCTTCTTCCAGCTGCT 3240
Db 3438 AAAACAGGAGTTTNGAAAAAACCTGAAATTGGCCAGTTTATTGCTTCTTCCAGCTGCT 3497
QY 3241 ACTGCTCCCGAACCACTGACCTGGCTCTGGGAGCTGATTCGATCCGGAAAAAGAAC 3300
Db 3498 ACTGCTCCCGAACCACTGACCTGGCTCTGGGAGCTGATTCGATCCGGAAAAAGAAC 3357
QY 3301 CCAGGTGATGATGAGGAGAGAACTGCAAGCTCAGAGGAAAAAGCCAGATAGGGTGG 3360
Db 3558 CCAGGTGATGATGAGGAGAGAACTGCAAGCTCAGAGGAAAAAGCCAGATAGGGTGG 3617
QY 3361 TTTCCAGCAATTATGTCAAACTTCTAAAGCCCGGAGCAAGCAAAATACCCCACTGAG 3420
Db 3618 TTTCCAGCAAAATTATGTCAAACTTCTAAAGCCCGGAGCAAGCAAAATACCCCACTGAG 3677
QY 3421 CTACCCAGAACCCGAGTCAGCAGCAGTGGCCAGGTGATGGGATGTAGGATTACACC 3480
Db 3678 CTACCCAGAACCCGAGTCAGCAGCAGTGGCCAGGTGATGGGATGTAGGATTACACC 3737
QY 3481 GCCCAGAAAGATGACGAATAGCTTTCAGCAAAAGGCCAGATCATCAAGCTCTCAACAG 3540
Db 3738 GCCCAGAAAGATGACGAATAGCTTTCAGCAAAAGGCCAGATCATCAAGCTCTCAACAG 3797
QY 3541 GAGGACCCGAGCTGTTGAAAGAGAGAGTCACTGAGGCAAGTTGGGCTTCTCCCATCAAT 3600
Db 3798 GAGGACCCGAGCTGTTGAAAGAGAGAGTCACTGAGGCAAGTTGGGCTTCTCCCATCAAT 3857
QY 3601 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCCAGCAATGA 3642
Db 3858 TATGTAAAGCTGACCAAGACATGAGACCCCAAGCCAGCAATGA 3899

```

RESULT 5
AAZ34572 standard; cDNA; 5195 BP.
ID AAZ34572 standard; cDNA; 5195 BP.

AAZ34572;
DT 01-FEB-2000 (first entry)
DE Human SH3D1A cDNA clone 11.
XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
XX megakaryocytic abnormality; myeloproliferative disorder;
XX platelet disorder; neural disorder; thrombocytopenia;
XX haematopoietic disorder; cognitive dysfunction; microcephaly;
XX liisencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
XX 88.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 239..3886
FT /tag= a
XX MO9953062-A2.

```

XX 21-OCT-1999.
PD 16-APR-1999; 99MO-US08371.
XX 16-APR-1999; 98US-0082007.
PR 16-APR-1998; 98US-0082007.
XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
PA Korenberg JR, Chen X;
PI WPI; 1999-633829/54.
DR P-PSDB; AAY32156.
XX Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukaemia -
PS Claim 2; Fig 10; 99p; English.
XX This is the nucleotide sequence of full-length cDNA (clone 11)
CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
CC that contributes to the development of platelets and the
CC pathogenesis of leukaemias, both in general and in particular those
CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
CC small candidate region for low platelets on chromosome 21.
CC Sequencing of 5 different sizes of cDNA clone from foetal brain
CC (see AAZ34570-74) suggests that at least 3 isoforms exist. The
CC invention provides methods for the diagnosis and treatment of
CC megakaryocytic abnormality, myeloproliferative disorder, platelet
CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
CC platelet disorder on chromosome 21, low platelets in deletion for
CC 21, association of gains in chromosome 21 with leukaemias, neural
CC abnormalities, dysfunctions and disorders including brain
CC malformations and corresponding cognitive dysfunctions, are also
CC microcephaly, liisencephaly, and colpocephaly. Methods are also
CC provided for: suppressing cells unable to regulate themselves;
CC screening for a somatic alteration in the SH3D1A gene; monitoring
CC the progress and adequacy of a treatment; monitoring tumour risk
CC progress or megakaryocytic abnormality, myeloproliferative disorder,
CC haematopoietic disorder, platelet disorder or leukaemia; and
CC treatment of a subject (including a prenatal subject) having
CC megakaryocytic abnormality, myeloproliferative disorder,
CC platelet disorder, leukaemia or neural disorder using a
CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 other;
SQ
Query Match 78.4%; Score 2856; DB 20; Length 5195;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 3167; Conservative 0; Mismatches 475; Indels 6; Gaps 2;
QY 1 ATGGCTCAGTTTCCCAACCTTTGGTGTAGCTGATGTCGGCCATACCTGTGAG 60
Db 239 ATGGCTCAGTTTCCCAACCTTTGGTGTAGCTGATGTCGGCCATACCTGTGAG 298
QY 61 GAAAGGGCCAGATGACGACGACCTTCTTAAAGCCGATGAGGGGATTTATTACT 120
Db 299 GAAAGGGCCAGATGACGACGACCTTCTTAAAGCCGATGAGGGGATTTATTACT 358
QY 121 GGTATCAAGCGAGAACTTTTTCATCTGGGTACTCTCACTGTTTACGACAA 180
Db 359 GGTATCAAGCGAGAACTTTTTCATCTGGGTACTCTCACTGTTTACGACAG 418
QY 181 ATATGGGCGCTTACGAGCATGATACGATGAGAGATGATCAAGTGAATTTTCCATA 240
Db 419 ATATGGGCGCTTACGAGCATGATACGATGAGAGATGATCAAGTGAATTTTCCATA 478
QY 241 GCCATGAAGCTTATCAATCAAGTCAAGATATACGCTCCCTCCACACTCCCTCCCT 300
Db 479 GCTATGAACCTTATCAATCAAGTCAAGATATACGCTCCCTCCACACTCCCTCCCT 538
QY 301 GTCATGAACGACCAAGTGGCTATTTCAGTGCACACGATTTGGTATGAGAGGATT 360

```


Db 539 GTCATGAACAGCAACAGTGTCTATTTCTAGCGCACAGCATTTGTGAGAGATATC 598
Qy 361 GGTAGATGCCCACTCAAGCTGTGCTCTGTGCCAATGGGCTCCATTCAGATTGTT 420
Db 599 GCCAGATGCCACCGCTTACAGCTGTGTCTCCAGTCCCAATGGGAATCCATTCAGATTGTT 658
Qy 421 GGAATGTCCACCCCTTAGTATCTTCTGTCCCTCCAGCAGCAGTGTGCTCCCTGGCTAAC 480
Db 659 GGAATGTCTCCAACCCCTAGTATCTTCTGTCCCAAGCAGCAGTGTGCTCCCTGGCTAAC 718
Qy 481 GGAGTCTCTCCCTCAATACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 540
Db 719 GGAGTCTCTCCCTGTATACACCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 778
Qy 541 AAGAGTCTCTCTCAAGCAGATTTGTGCTCAAGGCTCAATTAACCTAAGTTACGAAG 600
Db 779 AAGAGTCTCTCTTATGATGATCTGTGCTCAAGGCTCAATTAACCTAATTAACGAAG 838
Qy 601 GCACATCATTCGATGTGCGCAGCGCCCTCCAGCAGCAAGATGGGCTGTGCTCAGTCA 660
Db 839 GCACATCATTTGATGTGCGCAGGTGTCCACACAGTGGCAGAGTGGGCTGTCTCAGTCA 898
Qy 661 TCAAGGCTGAATATACAGGCAATTAATTAACAGCCACAGCAAAAATAAGTGAACATTA 720
Db 899 TCAAGACTGAATATACAGGCAATTAATTAATGATCATGACAAAATAAGTGAACATTA 958
Qy 721 AAGAGTCTCTCAAGCAGCAATTAATTAATGATCATGACAAAATAAGTGAACATTA 780
Db 959 AAGAGTCTCTCAAGCAGCAATTAATTAATGATCATGACAAAATAAGTGAACATTA 1018
Qy 781 TCAATATGGAATCTTCTGATCATGATCAAGATGAGAAAATACTGTCAGAGAAATTAATC 840
Db 1019 TCAATATGGAATCTTCTGATCATGATCAAGATGAGAAAATACTGTCAGAGAAATTAATC 1078
Qy 841 CTAGCTATGCACTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db 1079 CTAGCTATGCACTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1138
Qy 901 CCAGATATCATCT 960
Db 1139 CCAGATATCATCT 1198
Qy 961 AGCT 1020
Db 1199 AGCT 1258
Qy 1021 C---CAGAGAGAACTGCTGTGATCTTGAATTAAGATTAAGAGCGGAGAACTTCAAGCA 1077
Db 1259 CAATTAGAAAGAAATTAATCTGTAACTTTGAAGATTAAGAGCGGAGAACTTTGAAGCT 1318
Qy 1078 GGCAGTGTGAGCTGAGAAAGCGCCCAAGCGCTTTGGAGCAGCAGCGCAAGAGCAG 1137
Db 1319 GGCAGCTGTGAATCTGAGAAAGCGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAG 1378
Qy 1138 GAGCGTGTGCTGAGCTGAGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
Db 1379 GAGCGCTGTGCTGAGCTGAGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1438
Qy 1198 CAGAGGCGCAAGCGCAGCTGAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAG 1257
Db 1439 CAGAGGCGCAAGCGCAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1498
Qy 1258 CCGCAGCGAG 1317
Db 1499 CCGCAG 1558
Qy 1318 CTGGAAGGAG 1377
Db 1559 CTGGAAGGAG 1618
Qy 1378 AGGAG 1437

Db 1619 AGAAGCAAG 1678
Qy 1438 GAGTTAGAGCTCTGATTAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1679 GAGTTAGAGCTCTGATTAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1738
Qy 1498 TGTGAGCTGGCAACCGAG 1557
Db 1739 TGTGAGCTGGCAACCGAG 1798
Qy 1558 ATTGCTGAAGTACCCATCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 1799 ATTGCTGAAGTACCCATCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
Qy 1618 CTATTTCCAG 1677
Db 1859 CTATTTCCAG 1918
Qy 1678 CATAGAGACTCGCTCTTACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
Db 1919 CATAGAGATTCCTTGTACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
Qy 1738 CAGCTCCGAG 1797
Db 1979 CAGCTACAG 2038
Qy 1798 GATGTTTCAACAG 1857
Db 2039 GATGTTTCAACAG 2098
Qy 1858 AAGCAG 1917
Db 2099 AAGCAG 2158
Qy 1918 GAGTTAG 1977
Db 2159 GAGTTAG 2218
Qy 1978 CTGAG 2034
Db 2219 CTGAG 2278
Qy 2035 AGAGTGAAG 2094
Db 2279 AGAGTGAAG 2338
Qy 2095 ATGCAAG 2154
Db 2339 ATGCAAG 2398
Qy 2155 CAGGAG 2214
Db 2399 CAGGAG 2458
Qy 2215 AAGAGTGAATTAACAG 2274
Db 2459 AAGAGTGAATTAACAG 2518
Qy 2275 ATCCAG 2334
Db 2519 ATCCAG 2578
Qy 2335 GAG 2394
Db 2579 GAG 2638
Qy 2395 GAG 2454
Db 2639 GAG 2698
Qy 2455 AAGAGTGAATTAACAG 2514
Db 2699 AAGAGTGAATTAACAG 2758

QY 2515 ACCCCCAACAACCTGGGAGACTTTCAGTTCACAGTGGCCCAAGCCTCAAGAGAGCA 2574
 DB 2759 ACCCTTAATTAACCTGGCCGACTTCAGCTCACGTGGCCCAAGAGAGAGAGAGCA 2818
 QY 2575 GAAACGAGCAACCTGGGATACGTGGGCGGCTCAGCCTTCTCTGACCGTACTAGTGTGC 2634
 DB 2819 GAAACGAGTAACCTGGGATACGTGGGCGGCTCAGCCTTCTCTGACCGTACTAGTGTGC 2878
 QY 2635 CAGTTACGGCAGAGATCAGCCTTTACCCAGCCACAGCCACTGGCTCTCCCATCTCC 2694
 DB 2879 CAGTTACGGCAGAGATCAGCCTTTACCCAGCCAGCCACTGGCTCTCCCATCTCTCT 2938
 QY 2695 GTCCGGGGCCAGGGGAGAGAGGTTGAGAGGCTCAAGGCGCAAGCCCTGATCCCGGAGA 2754
 DB 2939 GTGCTAGGCGAGGGGAGAGAGGTTGAGAGGCTCAAGGCGCTATATCTTGGAGA 2998
 QY 2755 GCCAAAAAGACACACCTTAATTTTAAACAAGTGAAGTATCAGCCTTCTGGAGAG 2814
 DB 2999 GCCAAAAAGACACACCTTAATTTTAAACAAGTGAAGTATCAGCCTTCTGGAGAG 3058
 QY 2815 CAAAGCATGTGGTGTGGAGAGTTCAAGTCAAGAGGTTGGTCCCAAGTCTTAC 2874
 DB 3059 CAAAGCATGTGGTGTGGAGAGTTCAAGTCAAGAGGTTGGTCCCAAGTCTTAC 3118
 QY 2875 GTGAAACATCATTTTGGAGAGGCGGTAGGAATCAAGAGATGATACCTGGCCCTAGAA 2934
 DB 3119 GTGAAACATCATTTTGGAGAGGCGGTAGGAATCAAGAGATGATACCTGGCCCTAGAA 3178
 QY 2935 AGTCCGTAGTCTTAAGAGAGTGGCTTCCCGGCGGCAAGCCATTCCTCGAGAA 2994
 DB 3179 AGTCCGTAGTCTTAAGAGAGTGGCTTCCCGGCGGCAAGCCATTCCTCGAGAA 3238
 QY 2995 GAGTTATTTCCATGTACATACAGAGTTTGAAGCAAGAGATTTAACTTTAGAGAA 3054
 DB 3239 GAGTTATTTCCATGTACATACAGAGTTTGAAGCAAGAGATTTAACTTTAGAGAA 3298
 QY 3055 GGGGATGTATGTGTGTACCAAGAAAGTGTGCTGGTGAAGGGAAGGTTGGGCGAC 3114
 DB 3299 GGGGATGTATGTGTGTACCAAGAAAGTGTGCTGGTGAAGGGAAGGTTGGGCGAC 3358
 QY 3115 AAGTCGGAGTCTTCCCTTCTAATGATGAGGCTTAAAGATTCAAGAGGCTCTGGAAT 3174
 DB 3359 AAGTCGGAGTCTTCCCTTCTAATGATGAGGCTTAAAGATTCAAGAGGCTCTGGAAT 3418
 QY 3175 GCTGGGAAACAGGAGTTTGAAGAAAAAAGCTGAAATTTGCCAGTTATTTGCTTAC 3234
 DB 3419 GCTGGGAAACAGGAGTTTGAAGAAAAAAGCTGAAATTTGCCAGTTATTTGCTTAC 3478
 QY 3235 GCTGTACGTGTCGGGAAACAATCAACCTGGCTCCGGGCGAGTGAATCTGATCCGGAA 3294
 DB 3479 ACCGCAACGGGCGGAGAGCTCACTCTGCGCTGTCAGCTAATTTGATCCGAAAA 3538
 QY 3295 AAGAACCGAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCAGATA 3354
 DB 3539 AAGAACCGAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCAGATA 3598
 QY 3355 GGGTGTGTTCCAGCAATTTATGTCAAACTTTGAGCCCGGAAACAAGCAAAATCAACCA 3414
 DB 3599 GGGTGTGTTCCAGCAATTTATGTCAAAAGCTTTCAGCCCTGGAGCAAGCAAAATCACTCA 3658
 QY 3415 ACTGAGCTAACCAAGCCGAGTGAAGCAGAGAGTGTGCAAGTATGAGGAGTGAACAT 3474
 DB 3659 AAGAGCCCACTTAAGTCAACACATTTAGGCGAGTGTGCAAGTATGAGGAGTGAACAT 3718
 QY 3475 TACACCGCCAGAGAGATGACGATAGCTTCAAGCAAGGCGCAGATCATCAAGCTCTC 3534
 DB 3719 TACACCGCCAGAGATGACGATAGCTTCAAGCAAGGCGCAGATCATCAAGCTCTC 3778
 QY 3535 AACAGAGAGAGATCCGAGCTGTGTGAAGAGAGAGTCAAGTGGCAAGTTGGGCTTTCCCA 3594
 DB 3779 AACAGAGAGAGATCCGAGCTGTGTGAAGAGAGAGTCAAGTGGGCTTTCCCA 3838

QY 3595 TCCATTATGTAAAGCTGACACAGATGAGACCCAGCCAGCAATGA 3642-
 DB 3839 TCCAATTATGTGAAGCTGACACAGATGAGACCCAGCCAGCAATGA 3886
 RESULT 6
 AA234571
 ID AA234571 standard; cDNA; 5458 BP.
 XX
 AC AA234571;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A cDNA clone 21.
 XX
 KW SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KW s8.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 267..3929
 FT //tag= a
 XX
 EN WO953062-A2.
 XX
 PD 21-OCT-1999.
 XX
 PP 16-APR-1999; 99WO-US08371.
 XX
 PR 16-APR-1998; 98US-0082007.
 XX
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 XX
 PI Korenberg JR, Chen X;
 XX
 DR WP1; 1999-633829/54.
 XX
 P-PSDB; AAY32155.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX
 PS Claim 2; Fig 8; 99p; English.
 XX
 CC This is the nucleotide sequence of full-length cDNA (clone 21)
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,
 CC that contributes to the development of platelets and the
 CC pathogenesis of leukaemias, both in general and in particular those
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the
 CC small candidate region for low platelets on chromosome 21.
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The
 CC invention provides methods for the diagnosis and treatment of
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,
 CC platelet disorder on chromosome 21, low platelets in deletion for
 CC 21, association of gains in chromosome 21 with leukaemias, neural
 CC abnormalities, dysfunctions and disorders including brain
 CC malformations and corresponding cognitive dysfunctions,
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also
 CC provided for: suppressing cells unable to regulate themselves;
 CC screening for a somatic alteration in the SH3D1A gene; monitoring
 CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a

CC nucleic acid that expresses SHD1A or its antisense nucleic acid.
XX Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 other;

Query Match 77.7%; Score 2831; DB 20; Length 5458;
Best Local Similarity 86.5%; Pred. No. 0;
Matches 3167; Conservative 0; Mismatches 475; Indels 21; Gaps 3;

```
QY 1 ATGGCTCAGTTTCCCAACACCTTCCGGGTGATGTCGGGCGAATTAATGAGG 60
DB 267 ATGGCTCAGTTTCCCAACACCTTCCGGGTGATGTCGGGCGAATTAATGAGG 326
QY 61 GAAAGGCCAAGCATGATGAGAGTTCATAGTTTAAAGCAATATCTGGATTCAT 120
DB 327 GAAAGGCCAAGCATGATGAGAGTTCATAGTTTAAAGCAATATCTGGATTCAT 386
QY 121 GGTGATCAAGCGAAGAACTTTTTCATCTGGGTACCTGAGCCTGCTTACACA 180
DB 387 GGTGATCAAGCGAAGAACTTTTTCATCTGGGTACCTGAGCCTGCTTACACA 446
QY 181 ATATGGGCGCTAGCGAGATGAATAGATGAAGATGATGAATTTTCCAT 240
DB 447 ATATGGGCGCTAGCGAGATGAATAGATGAAGATGATGAATTTTCCAT 506
QY 241 GCCATGAAGCTTATCAAACTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 300
DB 507 GCATGAAGCTTATCAAACTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 566
QY 301 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACAGCATTTGGTATGAGGAG 360
DB 567 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACAGCATTTGGTATGAGGAG 626
QY 361 GGTAGAGCGACACATCAGCTGTTGCTCTGTCGCAATGGGCTCCATTCAGTT 420
DB 627 GGTAGAGCGACACATCAGCTGTTGCTCTGTCGCAATGGGCTCCATTCAGTT 486
QY 421 GGAATGTCACACCTTAGATCTTCTGTCCTCCAGCAGAGTGGCTCCCTGGCTAA 480
DB 687 GGAATGTCACACCTTAGATCTTCTGTCCTCCAGCAGAGTGGCTCCCTGGCTAA 746
QY 481 GGGGCTCTCCCTCATACAGCTTGGCTGCTGCTTTGGCGATCTGCAAGCAGATG 540
DB 747 GGGGCTCTCCCTCATACAGCTTGGCTGCTGCTTTGGCGATCTGCAAGCAGATG 806
QY 541 AAGAGTTCTTCTTCCAGCAGATCTGTCAGAGGCTCAATTAACCTAAGTTACGA 600
DB 807 AAGAGTTCTTCTTCCAGCAGATCTGTCAGAGGCTCAATTAACCTAAGTTACGA 866
QY 601 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGCAGCAAGTGGCTGCTCAGTCA 660
DB 867 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGCAGCAAGTGGCTGCTCAGTCA 926
QY 661 TCAGAGCTGAATATACAGGAGTTTCAACAGCCACAGCAAAAATAAGTGAAGCA 720
DB 927 TCAGAGCTGAATATACAGGAGTTTCAATATGTCATGACAAACTATAGTGAAG 986
QY 721 AAGAGTCCCGCAGCAGAACTATTTCTATGATCAAGTTTACCCAGAGCTCAGTGG 780
DB 987 AAGAGTCCCGCAGCAGAACTATTTCTATGATCAAGTTTACCCAGAGCTCAGTGG 1046
QY 781 TCATATGGAATCTTCTGATTCATGATCAAGTGAAGAACTCACTCAGAGAAATTAT 840
DB 1047 TCATATGGAATCTTCTGATTCATGATCAAGTGAAGAACTCACTCAGAGAAATTAT 1106
QY 841 CTAGCTATGACCTTAATGATGTTGCTGATCTGTCAGCAGCTGCGCCGCTGCT 900
DB 1107 CTAGCTATGACCTTAATGATGTTGCTGATCTGTCAGCAGCTGCGCCGCTGCT 1166
QY 901 CCGAATATCATCTCTCTCTTCCAGAGAGTTCGCTCGGCGAGTGGAGATGCCAT 960
DB 1167 CCGAATATCATCTCTCTCTTCCAGAGAGTTCGATCTGCGAGTGGATATCTGTCAT 1226
QY 961 AGCTTTCTCTGATGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGATGAGCAG 1020
```

```
DB 1227 AGCTCAACATCTGTAGATCAGAGGCTACAGAGGAACAGTTTGAAGATGAACAACA 1286
QY 1021 C---CAGAGAAAGAACTGCTGCTGATTCATTTGAAGATGAAGACGGGAGAACTTCAG 1077
DB 1287 CAATTAGAAAGAAATTAACCTTAACTGTTGAAGATGAAGACGGGAGAACTTTGAAC 1346
QY 1078 GCGAGTGTAGAGTGAAGAAAGCGCCGCAAGGCTCTTTGAGCAGCAGCGCAAGAGCAG 1137
DB 1347 GCGAGTGTAGAGTGAAGAAAGCGAGCTTCTTGAACAGCAGCGCAAGAGAGCAG 1406
QY 1138 GAGCGGTTGGCTCAGCTGAGCGCCGCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAG 1197
DB 1407 GAGCGGTTGGCTCAGCTGAGCGCGCGAGCAGAGAGAGAAAGAGAGAGAGCTGAGAG 1466
QY 1198 CAGAGGCCAAGCGGAGCTGAGAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAGAG 1257
DB 1467 CAGAGGCCAAGCGGAGCTGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1526
QY 1258 CCGCAGCGAGAGAGAGAGAGAGAGAGATCGAGAGAGCGGAGCGCAAAAGCGGAA 1317
DB 1527 CCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1586
QY 1318 CTGAAAGCGCAGCGCAACTTGAATGGAACGGAACCGAGACAGAGACTCTGATTCAG 1377
DB 1587 CTGAAAGCGCAGCGCAACTTGAATGGAACGGAATCGAAGCGCAAGACTTCAATTCAG 1646
QY 1378 AGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
DB 1647 AGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1706
QY 1438 GAGTTAGAGAGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
DB 1707 GAGTTAGAGAGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1766
QY 1498 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
DB 1767 TGTGAGCTGGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1826
QY 1558 ATTGCTGAAGATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
DB 1827 ATTGCTGAAGATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1886
QY 1618 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
DB 1887 CTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1946
QY 1678 CATAGAGACTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
DB 1947 CATAGAGACTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2006
QY 1738 CAGCTCCGGAGCAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1797
DB 2007 CAGCTCCGGAGCAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2066
QY 1798 GATGTTTCAACAACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
DB 2067 GATGTTTCAACAACAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2126
QY 1858 AAGCAGAGGCTCCCTGAGAGCAGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 1917
DB 2127 AAGCAGAGGCTCCCTGAGAGCAGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 2186
QY 1978 GATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1977
DB 2247 GATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
QY 2035 AGACTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
```

```

Db 2307 AAACGAAAGGAGAGAGTGTCAAAAAGAAAGATGGCGAGGAAAAAGGCAACAGGAA 2366
Qy 2095 ATGCAAGACAGAGAGTGGCTTTTCCATCCGATCGAGAGCCAGCTTAAGTGGCCACC 2154
Db 2367 GCACAAAGCAAGCTGGTGGCTTTTCCATCAACCAAGAAACGAGTAAAGCAGCTTTC 2426
Qy 2155 CAGGACCCCTGGTCTACCAAGAGAAAGCCGCTTACCATTTCTGCAAGAGAGTGA 2214
Db 2427 CAGGACCCCTGGTCTACCAAGAGAAAGTCCATTCATTTCTGCAAGAGAAATTA 2486
Qy 2215 AAAGTGATATATTCAGAGCGCTGTACCCCTTTGAATCCAGAAATCAGATGAGATCAC 2274
Db 2487 AAAGTGATATATTCAGAGCGCTGTACCCCTTTGAATCCAGAAATCAGATGAGATCAC 2546
Qy 2275 ATCCAGCCAGAGATATAGTCTAT-----GTTGATGAAAGCCAGACTGGA 2319
Db 2547 ATCCAGCCAGAGATATAGTCTATGTTAAAGGGAATGGTGAAGAAAGCCAACTGGA 2606
Qy 2320 GAGCCAGAGATGGCTTGAGAGAGAGCTGAAAGGAGAGAGCGGATGGTTCCCTGCAAACTAT 2379
Db 2607 GAGCCAGAGATGGCTTGAGAGAGATTTAAAGGAAAGAGAGGATGGTTCCCTGCAAACTAT 2666
Qy 2380 GCAGAAAGATTCAGAAAGATAGGTTCCACTCCAGCCAAACAGTACGATCTGACA 2439
Db 2667 GCAGAAAGATTCAGAAAGATAGGTTCCCGCTCCAGTGAACAGTACGATCTGACA 2726
Qy 2440 TCTGCCCTTCCGCTTCCGCTTCTGCTGAGACCCCTGCTCTTCTTCCAGTACCTCT 2499
Db 2727 TCTGCCCTTCCGCTTCCGCTTCTGCTGAGACCCCGCTTCTTCCAGTACCTCT 2786
Qy 2500 TCTGAGCCCTCCAGAACCCCAACAAGTGGGAGAGCTTCAATTCCAGTGGCCAGCAGC 2559
Db 2787 TCAAGAGCCCTCCAGAACCCCTTAATCTGAGGCGAGCTTCAAGTGGCCAGCAGC 2846
Qy 2560 TCAAAAGAGAGCCAGAAACCGAGCAATGAGATAGTGGCGGCTCAGCCTTCTCTGACC 2619
Db 2847 ACGAATGAGAAACAGAAACCGAGTAATCTGGATGATGGGCGAGCCAGCCCTCTCTGACC 2906
Qy 2620 GTACTAGTGTCTGCGCACTTACGAGAGATCAGCTTTACCCAGCCAGCAGCCTGGC 2679
Db 2907 GTTCCAAATGCGCGGAGCTTAAAGGAGAGTCCGCTTTATCCAGCCAGCAGCCTGGC 2966
Qy 2680 TCCCTCCCATCTCCCGTCTGGGCGAGGAGTGAAGAGGAGCTAACAGCGCAAGCC 2739
Db 2967 TCCCTCCCATCTCCCGTCTGGGCGAGGAGTGAAGAGGAGCTAACAGCTAACCC 3026
Qy 2740 CTGATATCCCTGGAGAGCCAAAGAACCACTTAAATTTTAAACAAAGTACGCTATC 2799
Db 3027 CTATATCTCTGGAGAGCCAAAGAACCACTTAAATTTTAAACAAAGTACGCTATC 3086
Qy 2800 ACCGTTCTGGAACAGCAAGACATGTGTGTTTGGAGAAATTCAAGTCAAGAGGTTGG 2859
Db 3087 ACCGTTCTGGAACAGCAAGACATGTGTGTTTGGAGAAATTCAAGTCAAGAGGTTGG 3146
Qy 2860 TTCCCAAGTCTTACGTGAAATCATTTCAAGGAGCCGTAAGGAATTCACAAGATGAT 2919
Db 3147 TTCCCAAGTCTTACGTGAAATCATTTCAAGGAGCCGTAAGGAATTCACAAGATGAT 3206
Qy 2920 ACTGGCCCTACTGAAGTCTGCTAGTCTAAAGAGATGCTTCCCGGCGCCAGGCA 2979
Db 3207 TCTGGTCTTCAAGAGTCTGCTAGTCTAAAGAGATGCTTCCCGGCGCCAGGCG 3266
Qy 2980 GCCATTCCCGAGAAAGATTATTTGCCATGTACATACGAGATTTCTGAGCAAGAGAT 3039
Db 3267 GTCTGTTGCGGAGAAATTTATTCATGTACATTCAGAGATTTCTGAGCAAGAGAT 3326
Qy 3040 TTAACCTTTCAGCAAGGAGATGATGAGTTACCAAGAAAGATGAGTGGGAGCG 3099
Db 3327 TTAACCTTTCAGCAAGGAGATGATGAGTTACCAAGAAAGATGAGTGGGAGCG 3386
Qy 3100 GGAACGATGGGAGCAAGATCCGAGTCTTCCCTTAACTATGTAGGCTTAAAGATTCA 3159
Db 3387 GGAACGATGGGAGCAAGATCCGAGTCTTCCCTTAACTATGTAGGCTTAAAGATTCA 3446

```

```

Qy 3160 GAGGCTCTGGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAATTGGCCAG 3219
Db 3447 GAGGCTCTGGAACTGCTGGGAAAAACAGGAGTTTAGAAAAAAACCTGAATTGGCCAG 3506
Qy 3220 GTTATGCTTCTCAAGCTGTACTGTGTCCGAAACACTCACTCTGCTCTGGGAGCTG 3279
Db 3507 GTTATGCTCTATACACCGCACCGGCCCGAGAGCTCACTCTGCTCTGGGAGCTG 3566
Qy 3280 ATTCTGATCCGAAAAAGAACCCAGTGTATGTGGAGAGAGAACTGCAAGCTCGAGG 3339
Db 3567 ATTCTGATCCGAAAAAGAACCCAGTGTATGTGGAGAGAGAACTGCAAGCACTGGG 3626
Qy 3340 AAAAGGCCAGATAGGATGGTTTCCAGCAAAATTAATGCTTAAAGCCCGGAGCA 3399
Db 3627 AAAAGGCCAGATAGGATGGTTTCCAGCTAATTAATGAAAGCTTAAAGCTCTGGAGC 3686
Qy 3400 AGCAAAATCAACCCCACTAGCTACCCAGACCCAGTGCAGCCAGAGTGTGCCAGTG 3459
Db 3687 AGCAAAATCACTCCAGAGAGCCACTAAGTCAACAGCATTAAGGCGAGTGTGCCAGTG 3746
Qy 3460 ATCGGATGTACGATTAACCGCCAGAGAGATGACGAACTAGCTTCAAGAAAGCCAG 3519
Db 3747 ATTGGATGTACGATTAACCGCCAGAGATGACGAACTAGGCTTCAACAGAGGCGAG 3806
Qy 3520 ATCATCAACGCTCCCAACAGAGAGAGCCGCACTGGTGAAGAGAAAGTCAAGGGGCA 3579
Db 3807 ATCATCAACGCTCCCAACAGAGAGAGCCCTGACTGGTGAAGAGAAAGTCAAGAGCA 3866
Qy 3580 GTTGGGCTCTTCCCATCAATTAATGTAAGCTGACACAGATGAGACCCAGCCAGCAA 3639
Db 3867 GTTGGGCTCTTCCCATCAATTAATGTAAGCTGACACAGATGAGACCCAGCCAGCAA 3926
Qy 3640 TGA 3642
Db 3927 TGA 3929

RESULT 7
AAS84763
ID AAS84763 standard; cDNA; 7435 BP.
XX AAS84763;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX DNA encoding novel human diagnostic protein #20567.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
XX W0200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX DR P-FSDB: ABG20576.
XX
XX New isolated polynucleotide and encoded polypeptides; useful in
XX PT diagnostics, forensics; gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -

```

Claim 1, SEQ ID No 20567, 103bp, English.

XX The invention relates to isolated polynucleotide (I) and
PS polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. A864197-A894564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX

Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 68.4%; Score 2491.6; DB 23; Length 7435;
Best Local Similarity 84.2%; Pred. No. 0;
Matches 3112; Conservative 0; Mismatches 529; Indels 53; Gaps 25;

QY 1 ATGGCTCAGTTTCCACACCTTTCGGTGGAGCTGGATGTCGGGCACTAATCTGGAG 60
DB 263 ATGGCTCAGTTTCCACACCTTTCGGTGGAGCTGGATGTCGGGCACTAATCTGGAG 322
QY 61 GAAAGGGCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
DB 323 GAAAGGGCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 382
QY 121 GGTGATCAAGCAGGAACTTTTTCATCTGGGTTACTGAGCTGCTTGAAGCA 180
DB 383 GGTGATCAAGCAGGAACTTTTTCATCTGGGTTACTGAGCTGCTTGAAGCA 442
QY 181 ATATGGGGCTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 443 ATATGGGGCTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
QY 241 GCCATGAACTTATCAACTGAAAGTACAGATATCACTCCCTCCACACTTCCCT 300
DB 503 GCCATGAACTTATCAACTGAAAGTACAGATATCACTCCCTCCACACTTCCCT 562
QY 301 GTTCATGAAACGACACAGTGGCTATTCAGTGCACACGACATTTGATAGAGGAT 360
DB 563 GTTCATGAAACGACACAGTGGCTATTCAGTGCACACGACATTTGATAGAGGAT 622
QY 361 GTTAGAGGCACTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
DB 623 GTTAGAGGCACTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
QY 420 TGGAAATGTCACACCTTAGTATCTTCTGCTCCAGAGAGAGTCCCTCCCTGGCTAA 479
DB 683 TGGAAATGTCACACCTTAGTATCTTCTGCTCCAGAGAGAGTCCCTCCCTGGCTAA 742
QY 480 CGGGGCTCTCCCGCATACAGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 539
DB 743 CGGGGCTCTCCCGCATACAGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 802
QY 540 AAGAGTTCTCTCTGAGAGATCTGCTGAGGAGTCAATTAACATTAAGTTACAGAA 599
DB 803 AAGAGTTCTCTCTGAGAGATCTGCTGAGGAGTCAATTAACATTAAGTTACAGAA 862
QY 600 GGCACATCATTCATGATGCTGCGACGCGCCCTCC-AGCAGCAGATGGCTGCTCACT 658

DB 863 GGCACATCATTCATGATGCTGCGACGCGCCCTCC-AGCAGCAGATGGCTGCTCACT 922
QY 659 CATCAAGCTGAAATATCA---GGCAGTTATTAACAGCCAGCAAAAC--TATGATGG 713
DB 923 CATCAAGCTGAAATATCAAGCAAAATATTAATCAATGATCAAAACCTAATGATGG 982
QY 714 ACACCT--AACAGTCCCGAGCAAGAACTATTCATGATCAATCAAGTTTACCCAGGCT 771
DB 983 ACACCTTAAACAGTCCCGAGCAAGAACTATTCATGATCAATCAAGTTTACCCAGGCT 1042
QY 772 CAGCTGGCTTCAATATGAAATCTTTCTGATCAATGATCAAGTGAAGAACTCACTGAGAA 831
DB 1043 CAGCTGGCTTCAATATGAAATCTTTCTGATCAATGATCAAGTGAAGAACTCACTGAGAA 1102
QY 832 GAATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891
DB 1103 GAATTTATCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1162
QY 892 GTCTGGCTCCAGAAATACATCCCTCTCTCAGAGAGTTCGCTCCGCGAGTGGAGT 951
DB 1163 GTCTGGCTCCAGAAATACATCCCTCTCTCAGAGAGTTCGCTCCGCGAGTGGAGT 1222
QY 952 TCCGTCATAGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1011
DB 1223 TCCGTCATAGCTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1282
QY 1012 GAGCGCAGC---CAGAGAAATCTGCTGATGATTTGAATTAAGAGCGGAGAAC 1068
DB 1283 GAGCGCAGC---CAGAGAAATCTGCTGATGATTTGAATTAAGAGCGGAGAAC 1342
QY 1069 TTGAGCGAGCGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1128
DB 1343 TTGAGCGAGCGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1402
QY 1129 AAGAGCAGAGAGCGGTTGCTGAGTGGAGCGCGCGAGCAGAGAGAGAAAGCGAGG 1188
DB 1403 AAGAGCAGAGAGCGGTTGCTGAGTGGAGCGCGCGAGCAGAGAGAGAAAGCGAGG 1462
QY 1189 CGCCAGAGCAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1248
DB 1463 CGCCAGAGCAGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1522
QY 1249 GAGCTGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1308
DB 1523 GAGCTGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1582
QY 1309 AAGCGGAACT--GGAAGGCGAGCGCAACTTGAATGGGAAACCGAGACAGGAACT 1367
DB 1583 AAGCGGAACT--GGAAGGCGAGCGCAACTTGAATGGGAAACCGAGACAGGAACT 1642
QY 1368 CCTGAATCAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1427
DB 1643 ACTTAATCAAGAAACAG 1702
QY 1428 TCTGAGATTGAGTTAGAAAGCTCTGAATGACAAAGAGATCTGAGAAAGAACTTCA 1487
DB 1703 TCTGAGATTGAGTTAGAAAGCTCTGAATGACAAAGAGATCTGAGAAAGAACTTCA 1762
QY 1488 GATATCAGAGTGTGATGTCAGTCCAGACCCAGAGGCAAAATTTGAGACAGCACTTGA 1547
DB 1763 GATATCAGAGTGTGATGTCAGTCCAGACCCAGAGGCAAAATTTGAGACAGCACTTGA 1822
QY 1548 AGAGCTAAGATTGCTG--AAATCACCACTTACAGCAGAGTT--GCAAGATCTCAGC 1603
DB 1823 AGAGCTAAGATTGCTG--AAATCACCACTTACAGCAGAGTT--GCAAGATCTCAGC 1882
QY 1604 AAA---TGCTTGAAGAGCTTATTCAGAGAAACATATCTGAGTACAGAGTTAAACAG 1660
DB 1883 AAAAGCTTTGAGAAACATTTATTCAGAGAAACATATCTGAGTACAGAGTTAAACAG 1942
QY 1661 TCCAGCAGAACTTTGATAGAGAGCTGCTTCTTACCTCAAAAG-AGCTTGAAGCA 1719
DB 1943 TCCAGCAGAACTTTGATAGAGAGCTGCTTCTTACCTCAAAAGAGCTTGAAGCA 2002

Key	Location/Qualifiers
CDS	208..3642
	/*tag= a
WO9953062-A2.	
21-OCT-1999.	
16-APR-1999;	99WO-US08371.
16-APR-1998;	98US-0082007.
(CEDA-) CEDARS SINAI HEALTH SYSTEM.	
Korenberg JR, Chen X;	
WPI, 1999-633829/54.	
P-PSDB; AAY32154.	

Nucleic acid from the human SH3D1A gene and its products, useful for the diagnosis and treatment of myeloproliferative disorders and leukaemia -

Claim 2, Fig 5, 99pp; English.

CC This is the nucleotide sequence of full-length cDNA corresponding
CC to a novel human SH3 gene, termed the SH3DA gene, that contributed
CC to the development of platelets and the pathogenesis of leukaemias
CC both in general and in particular those involving the
CC megakaryocytic lineage. The SH3DA gene maps to the small
CC candidate region for low platelets on chromosome 21. Sequencing
CC of 5 different sizes of cDNA clone (see AM234570-74) suggests that
CC at least 3 isoforms exist. The invention provides methods for the
CC diagnosis and treatment of megakaryocytic abnormality,
CC myeloproliferative disorder, platelet disorder, acute leukaemia,
CC neural disorders, thrombocytopenia, platelet disorder on
CC chromosome 21, low platelets in deletion for 21, association of
CC gains in chromosome 21 with leukaemias, neural abnormalities,
CC dysfunctions and disorders including brain malformations and
CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
CC and colpocephaly. Methods are also provided for: suppressing
CC cells unable to regulate themselves; screening for a somatic
CC alteration in the SH3DA gene; monitoring the progress and
CC adequacy of a treatment; monitoring tumour risk progress or
CC megakaryocytic abnormality, myeloproliferative disorder,
CC haematopoietic disorder, platelet disorder or leukaemia; and
CC treatment of a subject (including a prenatal subject) having
CC megakaryocytic abnormality, myeloproliferative disorder,
CC platelet disorder, leukaemia or neural disorder using a
CC nucleic acid that expresses SH3DA or its antisense nucleic acid.
XX
XX Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;

Query Match	Score	DB	Length
Best Local Similarity 81.1%;	2429.6;	20;	5199;
Matches 2960; Conservative	0; Pident	469; Indels	219; Gaps

Query	1	208	61	268	121	
ATGGCTCAGTTTCCACACACCTTTCGATGTAGCTGATGTCGTGGCCATAACTGTGAG	60	ATGGCTCAGTTTCCACACACCTTTCGATGTAGCTGATGTCGTGGCCATAACTGTGAG	267	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	120	
208	ATGGCTCAGTTTCCACACACCTTTCGATGTAGCTGATGTCGTGGCCATAACTGTGAG	267	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	120	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	327
61	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	120	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	327	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180
268	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	327	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180
121	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180	GAAGAGGCCAAGCATGACAGAGTTCCTTAGCCCTGAAGCCGATAGCGGATTTAACT	180

Dd	328	GGTGTACAGCTGAAACTTATTTTTCATCTGGGTACCTCAACCTGTTTAGCAG	387
Qy	181	ATATGGCGCTAGGCGACATGAAATACGATGAAGATGATCAAGTGAATTTTCCATA	240
Dd	388	ATATGGGCACTAGCTGACATGAATATATGATGGAAGATGATCAAGTGAATTTTCCATA	447
Qy	241	GCCATGAAGCTTATCAAACTGAAGCTACAGAGATATACGCTCCCTCCACATTTCCCT	300
Dd	448	GCTATGAACCTTATCAAACTGAAGCTACAGAGATATACGCTCCCTCCACATTTCCCT	507
Qy	301	GTCATGAAGCAAGCAACGATGGCTATTTCCAGTCCACACGAGATTTGATATGAGGAT	360
Dd	508	GTCATGAAGCAAGCAACGATGGCTATTTCCAGTCCACACGAGATTTGATATGAGGAT	567
Qy	361	GCTAGCATGCGACACTCAACAGCTGTTGCTCTGTGCCAATGGGCTCCATTCAGTGT	420
Dd	568	GCGAGCATGCCACCGCTTACAGCTGTGCTCCATGCGCAATGGGATTCATTCAGTGT	627
Qy	421	GGAATGTCTCACCCCTTAGTATCTTGTGCTCCAGACGAGTCCCTCCCTGGCTAAC	480
Dd	628	GGAATGTCTCAACCCCTTAGTATCTTGTGCTCCAGACGAGTCCCTCCCTGGCTAAC	687
Qy	481	GGGGCTCTCCCGCTCATACAGCTCTGCTGCGCTTTGGGATCTCTGACGCCATATGCGCA	540
Dd	688	GGGGCTCTCCCGCTGTATACAACTCTGCTGCAATTTGCTCATCTGCGACCAATGCGCA	747
Qy	541	AAGATTTCTTCTTCAGACATCTGCTGCTCAGGGCTCACAATTAACACTATGATTCAGAA	600
Dd	748	AAGATTTCTTCTTATGATATCTGCTGCTCAGGGCTCACACTTAACACTTAATTCAGAA	807
Qy	601	GCACATCATTTGATGTGCGCCAGCGCCCTTCAGACAGACAGATGGGCTGTGCTCAGTCA	660
Dd	808	GCAACGATCATTTGATGTGCGCCAGCGCTCCACCAATGGGAGATGGGCTTCTTCAGTCA	867
Qy	661	TCAAGGCTGAATACAGGCACTTATTCACAGCCACGACAAACTATGATGAGACCTTA	720
Dd	868	TCAAGGCTGAATACAGGCACTTATTCAMTGTATGACAAACTATGATGAGACACTTA	927
Qy	721	ACAGGCTCCCGGAGAGAACTATTCATGATCAATCAAGTTAACCCAGGCTCAGTGGCT	780
Dd	928	ACAGGCTCCCGGAGAGAACTATTCATGATGATCAAGTTAACCCAGGCTCAGTGGCT	987
Qy	781	TCAATATGGAATCTTCTGACATTTGATCAAGATGGAAGAACTCAGCTGACGAAGAAATTA	840
Dd	988	TCAATATGGAATCTTCTGACATTTGATCAAGATGGAAGAACTTAACGACGAAGAAATTA	1047
Qy	841	CTAGCTATGCACTTAATTTGATGTGCCATGTCTGGTCAGCCATGCGCGCCGTCTGCT	900
Dd	1048	CTAGCTATGCACTTAATTTGATGTGGCTATGTCTGGCAACCACTGCTGCTGCT	1107
Qy	901	CCAGATATCAATCCCTCCTCTCCAGAAAGTTGCTCCCGGCAATGGGATGTCGCTATA	960
Dd	1108	CCAGATATCAATCCCTCTTCTTGAAGAGTTGATCTGGCAATGGGATGTCGCTATA	1167
Qy	961	AGCTCTTCTTCTGATGATCAGAGGCTCCTGAGAGCGCTGTGAGAGATGAGCAGCAG	1020
Dd	1168	AGCTCAACATCTGTATGATCAGAGGCTCCTGAGAGCAACGTTTATGAAATGAAACACAA	1227
Qy	1021	C---CAGAGAAAGAACTGCTCTGTGACATTTGAAGATTAAGAGCGGAGAACTTGAAGCA	1077
Dd	1228	CAATTAAGAAAGAAATTAACCTGTAACTTTGAAGATTAAGAGCGGAGAACTTGAAGCT	1287
Qy	1078	GGCAGTGTGAGCTGAGAAAGCGCGGCAAGCGCTCTTGAAGCAGAGCGGCAAGAGCAG	1137
Dd	1288	GGCAGCTGTGAGCTGAGAAAGCGAGGCAAGCGCTCTTGAAGCAGAGCGGCAAGAGCAG	1347
Qy	1138	GAGCGGTTGGCTCAGCTGAGCGCGCGCAGCAGAGAGAGAAAGAGCGGAGCGCCAGAG	1197
Dd	1348	GAGCGGCTGGCGCAGCTGAGCGCGCGCAGCAGAGAGAGAGAGAGAGCGGCGCCAGAG	1407
Qy	1198	CAGAGGCGCAAGCGGCACTGAGCTGAGAGAGAGCTGAGAGAGAGCGGAGACTGAG	1257

Db	1408	CAAGAGCGCAAAAAGCACTGGAATCTGGAAAGCAACTGGAAAAGCAAGCAGCGGGAGCTGAGA	1467
Qy	1258	CGGACGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGGCGCGCAAAACGGGAA	1317
Db	1468	CGGCAAGAGAGAGAGAGAGAGAGAGAAATTTGAGAGCGAGAGGCTGCAAAAACGGGAA	1527
Qy	1318	CTGGAAAAGCAGCCGCAACTTGAATGGGAAAGGAAACGGGAGACAGGAACCTCTGAATAG	1377
Db	1528	CTTGAAAAGGCAACGACACTTGAGTGGGAAACGGAATCCAAAGCAAGAACTCTAATATAA	1587
Qy	1378	AGGAACAAGAGAGAGAGAGGAGACCGTGTCTGGAAGCGAAGAGAGAAACCTGGAAGTTT	1437
Db	1588	AGAAACAAGAAACAAGAGAGACATAGTGTTCAGAAAGCAAGAAATAACCTTGGAAATT	1647
Qy	1438	GAGTTAGAAAGCTCTGAATGACAAAAAGCATCAGTAAGAGAAAATTCAAGATATCAGG	1497
Db	1648	GAAATTAGAAAGCTCTAATGATTAATAAGCATCAACTAGAGAGGAAACCTCAAGATATCAGA	1707
Qy	1498	TGTGCACTGGCAACCCAGAGCCAGAAATTTGAGAGCAAGAACTAGTAGAGCTAAGA	1557
Db	1708	TGTGCAATTGACCAACCCAAAGCCAAATAATTGAGACAAACAAATCTAGAGAGTTGAGA	1767
Qy	1558	ATTGCTGAATTCACCCCACTTACAGAGAGATTGACAGAAATCTCAGCAAACTCTTGAAGA	1617
Db	1768	ATTGCGGAAATCAACCCACTCTACAGAACAAATTAAGAGAACTCAGCAAAATCTTGAAGA	1827
Qy	1618	CTTATATTCAGAGAAACAGATATCTCAGTACCGAGTTAAAAACAAGTCACAGCAACAGTTTG	1677
Db	1828	CTTATATTCAGAGAAACAGATATCTCAATGACCAATTAAAAACAAGTTCAAGCAACAGTTTG	1887
Qy	1678	CATAGAGACTGGCTTCTTAACCTCAAAAGAGCTTTGAAAGCAAAAGAGCTGGCCCGGAG	1737
Db	1888	CACAGAGATTCACCTTGTTACACTTAAAAAGAGCTTTGAAGCAAAAGAACTAGCTGGGAG	1947
Qy	1738	CAGCTCCGGAGAGAGCTGAGAGAGGTGGAGAGAGACAGAGTCAAAAGCTCAGAGATT	1797
Db	1948	CACCTACAGAGACCACTGATGAAATGAGAGAAAGAACTTAGTCAAACTACAGAGATT	2007
Qy	1798	GATGTTTTCAACAACAGCTGAGAGAACTGAGAGATACATAGCAACAGCAACTCCAG	1857
Db	2008	GATATTTTCAATTAATCAGCTGAGAGAACTTAAAGAAATATCAATTAAGCAACACTCCAG	2067
Qy	1858	AAGCAGAGAGTCCCTGAGAGCAGCGGACTGAAAGCAAAAGCAGAGAGAGAAAGCCTG	1917
Db	2068	AAGCAAAAGTCCATGAGAGCTGAAAGCACTGAAACGAAAGAAACAAGAACGAAGATCTA	2127
Qy	1918	GAGTTAGAGAAAGCAAAAGAGAGGCTCAGAGACGAGTTCAGAAAGGGAACAAGCAATGG	1977
Db	2128	GAAATTTAGAAAACAAAAGAAAGAAAGCCCAAGACAGAGCTCAGAAAGGGAACAAGAGTG	2187
Qy	1978	CTGAGACATGTGCACAGCAGAGAG--GAGAGGCAACGCCCCCGGAAACCCACAGAGAGAGC	2034
Db	2188	CTGAGAGCATGTGCACAGAGAGAGAGAGATCAGAGAACCAAGAAAATCTCAAGAGAGGAA	2247
Qy	2035	AGACTGAAGAGGAGAGACAGTGTCAAGAAAGAGGCGGAAAGAGAGCCAAAGCCGGA	2094
Db	2248	AAACTGAAAAGGAGAGAGAGTGTCAAAAAGAAAGATGCGCAGGAAAAAGCAACAGGAA	2307
Qy	2095	ATGCAAGCAAGCAGAGTCCGTTTTTCATCCGATCAAGAGCCAGGCTAACTGGCAAC	2154
Db	2308	GCACAAAGCAAGCTGGGTCCGTTTTTCATCAACCAACAAAGAACAGCTAACAGCTGTCTC	2367
Qy	2155	CAGGCAACCTGGTCTACCAACAGAGAAAGCCCGCTTACATTTCTGCACAGAGAGTGA	2214
Db	2368	CAGGCAACCTGGTCTACCTGCAGAGAAAAGTCCACTTACCATTTCTGCACAGAGAAAATGA	2427
Qy	2215	AAAGTGTATATTACCGAGCGGTGATCCCTTTTGAATCAAGAGTCAAGATGAATCAC	2274
Db	2428	AAAGTGTGTATTACCGGCACTGTACCCCTTTGAATCAAGAGCAATGAATCACT	2487
Qy	2275	ATCCAGCCAGAGATATAGTCAATGATGGATGAAGCCAGACTGGAGAGCCAGATGGCTT	2334
Db	2488	ATCCAGCCAGAGATATAGTCAATGATGGATGAAGCCAAACTGGAGAAACCGGCTGGGCTT	2547

QY	2235	GGAGAGAGCGTGAAGAGGAAAGACGGGATGGTTCCTGCAAACTATGACGAAAGAAATATCCA	2394
Db	2548	GGAGAGAGAAATTTAAAGAGAAAGACGGGGTGTTCCCTGCAACTATGACGAGAAATATCCA	2607
QY	2395	GAAGATGAGGTTCCTCACTCCAGCCAAACAGATGACCGATCTGCATCTGCCCTGCCCC	2454
Db	2608	GAAGATGAGGTTCCTCCGCTCCAGTGAAACAGATGACTGATTTCAACATCTGCCCTGCCCC	2667
QY	2455	AAATGGCTCTGGCTGAGACCCCTGCTCTTTGGCAGTGAACCTTTCTGAGCCCTCCACA	2514
Db	2668	AAATGGCTCTTGGCTGTGAGACCCCTGGCTTTTGGCACTAACCTTTCAAGACCTTCCACG	2727
QY	2515	ACCCCAACAACTGGGAGACTTCAGTTCTCACTGGGCCAGCAGCTCAAAAGAAAGCA	2574
Db	2728	ACCCCTAATACTGGCGCACTTCAGCTCCACGTGGGCCACAGACAGAAATGAGAAACCA	2787
QY	2575	GAAGACGACAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCGTACCTATGTCTGGC	2634
Db	2788	GAAGACGATTAATCTGGGATGACATGGGCGAGCCAGCCCTCTCACCGTTCCAAGTGGCGGC	2847
QY	2635	CAGTTAGGGACGAGATCAGCCTTTAACCCCAAGCCACAGCACTGGCTCTCCCACTTCCC	2694
Db	2848	CAGTTAAGGACGAGAGTCCGCTTTTACTCAGCCAGGCACTGGCTCTCTCCCGTCTCT	2907
QY	2695	GTCTGGGCGCAGGGGTGAAGAAAGGTGAAGGGCTACAGACGCAAGCCCTGTATCCCTGAGA	2754
Db	2908	GTGCTAGGCGCAGGGGTGAAGAAAGGTGAGGGGCTAACACTCAAGCCCTATATCTTTGAGA	2967
QY	2755	GCCAAAAAGACAAACCACTTAATTTTAAACAAAGTACGTCACACCTTCTGGAACAG	2814
Db	2968	GCCAAAAAGACAAACCACTTAATTTTAAACAAATATATGTCACTACCGTCTGTGAACAG	3027
QY	2815	CAAGACATGTGTGTGTGTGGAGAGTTCAAGGTGAGAGGGTGTTCCTCCAACTCTTAC	2874
Db	3028	CAAGACATGTGTGTGTGTGGAGAGTTCAAGGTGAGAGGGTGTTCCTCCAACTCTTAC	3087
QY	2875	GTGAACTCATTTTCAGGGCCCGTAAAGAAATCCACAGCATCGATCTGGCCCTTACTGAA	2934
Db	3088	GTGAACTCATTTTCAGGGCCCGCTAAGGAAGTCTAACAGCATGTGATTCGTGTTCTTCAGAG	3147
QY	2935	AGTCTCTGTGTCTAAAGAGAGTGGGCTTCCCGGCGCCCAAGCCAGACATTTCCCGAGAA	2994
Db	3148	AGTCTCTGTGTCTTAAAGAGAGTGGGCTTCTTCAGACAGCCAGCGGTGTTTCGGAGAA	3207
QY	2995	GAGTTTATTTGCAATGTACACATACGAGAGTTCTTGAGCAAGAGATTTAACTTTACAGCA	3054
Db	3208	-----	3207
QY	3055	GGGATGTGATTTGTGTTCACAGAAAGATGTGTACTGTGTGACCGGAAACGTGTGGCGAC	3114
Db	3208	-----	3207
QY	3115	AAGTCCGAGACTTCCCTTCTAATCTATGTGAGGCTTAAAGATTCAGAGGGCTCTGAACT	3174
Db	3208	-----	3207
QY	3175	GCTGGGAAAAACAGGAGCTTTAGAAAAAAAACCTGAAATTGGCCAGGTTATTCTTCTAC	3234
Db	3208	-----GAAATTGGCCAGGTTATTCTTCTTAC	3234
QY	3235	GCTGCTATCTGTGTCCCGAACAACATCAGCCTGGCTCTGTGGGACGTGATTCGATCCGGAAA	3294
Db	3235	ACGCGCACCGGCCCCGAGCACTACTCTGCCCCCTGTGACGTGATTTTGTATCCGAAA	3294
QY	3395	AAGAACCCAGGTGATGTGTGGAGAGAACTGCAAGCTCGAGGGAAAAAGCGCAGATA	3354
Db	3295	AAGAACCCAGGTGATGTGTGGAGAGAGAGCTGCAACACGTGGGAAAAAGCGCAGATA	3354
QY	3355	GGGTGTGTTCCAGCAATATATGTCAAACCTTCTAAGCCCCCGGAACAACCAAAATACACCCA	3414
Db	3355	GGGTGTGTTCCAGCAATATATGTCAAACCTTCTAAGCCCCCTGGAGAGACCAAAATACCTCCA	3414

QY 3415 ACTGAGCTACCCAGACCCGAGTGCAGCCAGCAGTGTCCAGTATCCGAGTATGATCATGAT 3474
DB 3415 ACAGAGCCACCTTAAGTCAACAGATTAAGGCGAGTGTGCAGTGTATGGATGTACGAC 3474
QY 3475 TACACCGCCCAAGACATACGAACTAGCTTTCAGCAAAAGCCAGATCATCAACGTCTTC 3534
DB 3475 TACACCGCCCAAGATGACATGATGCTGCGCTTTCACAAAGGCGCAGATCATCAACGTCTTC 3534
QY 3535 AACCAAGAGAGACCCGAGCTGTGGAAGAGAAAGTACGAGGCGGCTCTCCCA 3594
DB 3535 AACCAAGAGAGACCTGTAGCTGTGGAAGAGAAAGTCAATGACAAAGTGGGCTCTCCCA 3594
QY 3595 TCCAAATTATGTAAGCTGCACACAGCATGACCCGACCCAGCAATGA 3642
DB 3595 TCCAAATTATGTAAGCTGCACACAGCATGACCCGACCCAGCAATGA 3642
RESULT 9
ABK43498
ID ABK43498 standard; cDNA; 3319 BP.
AC ABK43498;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #78.
XX
KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KM hyperproliferative disorder; neoplasia; cardiovascular disorder;
KM cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KM adenocarcinoma; reproductive system disorder; testicular feminisation;
KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KM respiratory disorder; renal disorder; kidney failure; blood disorder;
KM myocardial infarction; wound healing; cell proliferation; skin aging;
KM food additive; food preservative; gene therapy; gene; ss.
OS Homo sapiens.
XX
XX WO200155318-A2.
FN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209466.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244747.
PR 08-NOV-2000; 2000US-0244747.
PR 08-NOV-2000; 2000US-0244776.
PR 08-NOV-2000; 2000US-0244778.
PR 08-NOV-2000; 2000US-024523.
PR 08-NOV-2000; 2000US-024524.
PR 08-NOV-2000; 2000US-024525.
PR 08-NOV-2000; 2000US-024526.
PR 08-NOV-2000; 2000US-024527.
PR 08-NOV-2000; 2000US-024528.
PR 08-NOV-2000; 2000US-024532.

PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR MPI; 2001-581633/65.
 DR P-PSDB; AAU87168.
 XX
 PT New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -
 XX
 PS Claim 1; SEQ ID No 88; 837pp; English.
 XX
 CC The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (II) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiodystrophies, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 56.1%; Score 2043.8; DB 23; Length 3319;
 Best Local Similarity 85.2%; Pred. No. 0;
 Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;
 QY 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATCTGGGCGCATTACTGGAG 60
 Db 159 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATCTGGGCGCATTACTGGAG 218
 QY 61 GAAAGGGCCAGCATGACCGAGTTCCTTAAAGCCGATGAGGCGATTAATTAAT 120
 Db 219 GAAAGAGCGAAGCATGATGACAGTTCCTTAAAGCCGATGAGGCGATTAATTAAT 278
 QY 121 GGTATCAAGCAGGAACTTTTTCCTGATCTGGGTTTAACTGAGCTGTTAGCAAA 180
 Db 279 GGTATCAAGCAGGAACTTTTTCCTGATCTGGGTTTAACTGAGCTGTTAGCAAG 338
 QY 181 ATATGGGCGCTAGCGCATGATTAATGATGAGAGGATGATCAAGTGAATTTCCATA 240
 Db 339 ATATGGGCGCTAGCGCATGATTAATGATGAGAGGATGATCAAGTGAATTTCCATA 398
 QY 241 GCGATGAAGCTTATCAAACTGAGCTAATGAGATATGAGTATGAGAGGATTT 300
 Db 399 GCTATGAACTTATCAAACTGAGCTAATGAGATATGAGTATGAGAGGATTT 458
 QY 301 GTCATGAAAGCAAGCAACGAGTGTATTTCCAGTGCACACGATTTGGTATAGAGGAT 360
 Db 459 GTCATGAAAGCAAGCAACGAGTGTATTTCCAGTGCACACGATTTGGTATAGAGGAT 518
 QY 361 GCTAGATGCGCACCATCTACAGCTGTGCTCTGTGTGCAATGGGCTCATTCAGTTGT 420
 Db 519 GCGAGATGCGCACCGCTTACAGCTGTGCTCTGTGTGCAATGGGCTCATTCAGTTGT 578
 QY 421 GGAATGTCTCCACCTTATGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 Db 579 GGAATGTCTCCACCTTATGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
 QY 481 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 Db 639 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698
 QY 541 AAGAGTTCTTCTTACAGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Db 699 AAGAGTTCTTCTTACAGCAATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758
 QY 601 GCAATATCTTCATGTCGCGAGCGCCCTCCAGCAGCAAGAGGCTGTGCTGCTGCTGCT 660
 Db 759 GCACAGTCATTTGATGTCGCGAGCGCCCTCCAGCAGCAAGAGGCTGTGCTGCTGCTGCT 818
 QY 661 TCAAGGCTGAAATACAGGAGTATTCAGAGCCAGCAAACTATGATGAGGACACTTGA 720
 Db 819 TCAAGGCTGAAATACAGGAGTATTCAGAGCCAGCAAACTATGATGAGGACACTTGA 878
 QY 721 AC-----AGGTCCCGAGGAGAACTATTTCTATGATCAATCAAGTTTACCCGAG 768
 Db 879 ACAGGTTCTGTTTAAAGTCCCAAGCAAGAACTATTTCTATGATCAATCAAGTTTACCCGAG 938
 QY 769 GCTAGCTGGCTTAAATATGAAATCTTTCGATCTGATGATGAAATGAAATGAAATGAAAT 828
 Db 939 GCTAGCTGGCTTAAATATGAAATCTTTCGATCTGATGATGAAATGAAATGAAATGAAAT 998
 QY 829 GAAAGATTTATCTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGAT 888
 Db 999 GAAAGATTTATCTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGAT 1058
 QY 889 CCGGCTGCTCCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 948
 Db 1059 CCGGCTGCTCCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1118
 QY 949 ATGTCTGCTATTAATCTTCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1008
 Db 1119 ATATCTGCTATTAATCTTCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1178

OY	1009	GATGAGCAGCAGC---	CA	GAGAAGAACTGCGTGTGACATTGTGAAGTAATAAGCGGGAG	1005
Db	1179	GATGAAACAACAATTA	GA	AAATTAACCTTGAACTTTGAAGATAAAGCGGGAG	1238
OY	1066	AACTTCGAGCGAGCAGTGTGAGCTGTGAGAAAGCGCCCAAGCGCTTTGTGAGCAGAG	1125		
Db	1239	AACTTTGAACGTGGCACTTGGA	CT	TGAGAAACGAAGCAAGCTCTCTTGAAACAGAG	1298
OY	1126	CGCAAAAGCAGAGACGGGTTGGCTCAGCTTGAGCGCGCTCGAGCAGAGAGGAAAGCGG	1185		
Db	1299	CGAAAGAGAGCAGAGACGCGCTGGCCAGCTGGAGCGGGCGAGACAGAGAGGAAAGGAGCT	1358		
OY	1186	GAGCGCCAGAGCAGAGAGGCCAAGGGGCAAGCTGGAGCTGGAGAAAGACTGTGAAAGAG	1245		
Db	1359	GAGCGCCAGAGAGCAAGACCGCAAAAGACAACTGGAACTGGAGAGACCACTGAAAAGCAG	1418		
OY	1246	CGGAGCTGAGCGGCAAGCGGACCGAGAGGAGAGAGAGAGATGAGAGGCGCGAGGCGC	1305		
Db	1419	CGGAGCTTGAAGACGGCAGAGAGAGAGAGAGAGAGAAATTTGAGAGGCGAGAGGCT	1478		
OY	1306	GCAAAACGGGAACTGGAAAGGCAAGCAACTTGAATGGGAAACGGAAACGGAGACAGGA	1365		
Db	1479	GCAAAACGGGAACTTGAAGGCAACGAACTTGATGGGAAACGGAAATCGAAAGGCAAGAA	1538		
OY	1366	CTCCTGAATCAGAGGAACAAGAGCAGAGAGGCAACCGTGCTCTGAAGGCAAGAGCAG	1425		
Db	1539	CTACTTAATCAAGAGAAACAAAGAACAAAGAGACATAGTTGTACTGAAAGCAAGAAAG	1598		
OY	1426	ACTCTGAGTTTGAAGTTAGAAAGCTCTGAATGACAAAAAGCACTACGCTGAAGAGAACTT	1485		
Db	1599	ACTTTGGAATTTGAATTTGAAGCTTAATGTATTAAGATCACTTGAAGGGAACTT	1658		
OY	1466	CAGATATCAGGTGTGACTGGCAACCCAGAGCGCAAAATTGAGAGCAGCAACAAGTCT	1545		
Db	1659	CAAGTATCAGATGTGCAATTGCAACCCAAAGGCAAGAAATTGAGAGCAGCAACAATCT	1718		
OY	1546	AGAGAGCTTAAGATTGCTGAATTCACCACTTACAGCAGCTTGACAGAAATTCACGAA	1605		
Db	1719	AGAGAGTTGAAGATTGGCCGAAATCACCATCTACAGCAACAATTACAGAAATCTCAGCA	1778		
OY	1606	ATGCTGTGAAGACTTAATCCAGAGAAACAGATCTACGTGACGAGTTAAACAAGTCCAG	1665		
Db	1779	ATGCTGTGAAGACTTAATCCAGAGAAACAGATCTCAATGACCAATTAAACAAGTTGAG	1838		
OY	1666	CAGAACAGTTTGCAATAGAGACTCGCTTCTTAACCCCAAAAGAGCTTTGAAGCAAGAG	1725		
Db	1839	CAGAACAGTTTGCAATAGAGATTCACTTGTTCACCTTAABAAGCCTTGAAGCAAAAGAA	1898		
OY	1726	CTGGCCCCGCGCAGACGCTCCGGAGCAGCTTGACGAGGTGAGAGAGACCAAGTCAAG	1785		
Db	1899	CTAGCTCCGGCAGACCTTACGAGACCMACTGATGAAGTGAGAAAGAACTAGATCAAAA	1958		
OY	1786	CTGCAAGAGATTGAATGTTTTCACAACACAGCTGAAGGAACTGAGAGAGATCATAGCAA	1845		
Db	1959	CTACGAGGATTGATATTTTCAATATACGCTGAAGGAACTTGAAGAAATTCACAAATAG	2018		
OY	1846	CAGCAATCCAGAAAGCAGAGGTCCTCTGAGGCAAGCGCACTGAAGCAAGAAAGCAGAG	1905		
Db	2019	CAACAACCTCCAGAGACAAAGATCCATGAGGCTGAACGACTGAAACAGAAAGAACAGAA	2078		
OY	1906	AGAAAGACCTTGAGATTGAGAGCAAAAGAGACGCTCAGAGACGATTTCAAGAAAG	1965		
Db	2079	CGAAAGATCATAGATTTAGAAAAACAAAAAGAAAGAGCCCAAGACGAGCTCAAGAAAG	2138		
OY	1966	GACAAAGCAATGGCTGAGACATGTGACAGAGAG---GAGCAGCACGCCCCCGGAAACC	2022		
Db	2139	GAACAAGCAGTGCTGAGACATGTGACAGAGAGACGAGCATCGAGACCAAGAAATCTC	2198		
OY	2023	CACGAGAGAGCAGACTGAAGAGGAGAACAGTGTCAAGAGAAAGAGAGCGCGAAGAGAG	2082		
Db	2199	CACGAGAGAGAAAACTGAAAAAGGAGAGAGTGTCAAAAAGAGAGATGGCGAGAGAAAA	2258		
OY	2083	GCCAAGCCGGAATTCAGAACAGACAGAGTGGCTTTTCCATCCGCACTCAGAGACGACT	2142		

Db	2259	GGCAAAACGAGAGACAAAGCAAACTGAGTGGCTTTTCCATCAACCAAGAACAGCT	2318
Qy	2143	AAGCTGGCCACCAGGACCCCTGTGTTACACAGAGAAAGGCGCTTACCATTTCTGCA	2202
Db	2319	AAGCCAGCTGTCCAGGACACCTGTGTCCACTGCGAGAAAAGGTCCACTTACATTTCTGCA	2378
Qy	2203	CAGGAGAGTGTAAAGTGTATATATTACCGAGCGCTGTACCCCTTTGAATCCAGAGTCA	2262
Db	2379	CAGGAAAATGTAAAGTGTGTATTACCGGGCACTGTACCCTTTGAATCCAGAGCAT	2438
Qy	2263	GATGAGATCACCATCCAGCCAGGAGATATAGTCAATGTTGATGAAAGCCAGACTGAGAG	2322
Db	2439	GATGAAAATCACTATACCGCCAGGAGACATAGTCAATGATGAAAGCCAAACTGAGAA	2498
Qy	2323	CCAGGATGTGCTTGGAGAGAGCTGAAAAGGGAACGGGATGTTCCTTCGAAACTATGCA	2382
Db	2499	CCCGGCTGTGCTTGGAGAGAGATTTAAAGGAAAGCAGGGTGTTCCTTCGAAACTATGCA	2558
Qy	2383	GAAAAAGATTCCAGAAAAATGAGTTCCTCACTCCAGCCAAACAGAGACCGATCTGACATCT	2442
Db	2559	GAGAAAATCCCAAGAAAATGAGTTCCTCGCTCCAGTGAACCAAGTGAATCAACATCT	2618
Qy	2443	GCCCCCTGCCCCCAAACTGTGCTCTGCGTGAAGACCCCTGTCTCTTTGCCAGTGA	2502
Db	2619	GCCCCCTGCCCCCAAACTGTGCTCTGCGTGAAGACCCCGCCCTTTGGCAGTAACTCTTCA	2678
Qy	2503	GAGCCCTTCACAAACCCCAACAACTGGGACAGACTTCAGTTCCACGTGGCCAGAGCTCA	2562
Db	2679	GAGCCCTTCACAGACCCCTTATATATCTGGGCGGACCTTCAGCTTCACGTGGCCACAGCACG	2738
Qy	2563	AACGAGAGCCAGAAACGAGCAAACTGGGATACGTGGGCGGCTCAAGCTTCTCTGACCGTA	2622
Db	2739	AATGAGAAACCCAGAAACGAGATTAATCTGGGATGCAATGGGACGCGCCCTCTCTCA	2798
Qy	2623	CTTGTGTGTGGCCAGTTACGCGCAGAGATCAAGCTTTTACCCAGCCACAGCCACTGTGCTCC	2682
Db	2799	CCAAGTCCCGGCCAGTTAAGCAGAGGTTCGCGCTTACTCCAGCCACAGCCACTGTGCTCC	2858
Qy	2683	TCCCATCTCCCGTCCGTCGCGGCGCAGGNGA	2711
Db	2859	TCCCGTCTCTGTGTAGCGCAGCTTGA	2887
RESULT 10			
AA163825			
ID	AA163825 standard; cDNA; 3466 BP.		
XX	AA163825;		
XX	AC		
XX	22-OCT-2001 (first entry)		
DT			
XX	Human polynucleotide SEQ ID NO 33.		
DE			
XX	Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;		
KW	cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;		
KW	fungicide; ophthalmologic; cytosolic; immunosuppressive; nootropic;		
KW	neuroprotective; antiallergic; hepatotropic; antidiabetic;		
KW	antiinflammatory; antitumor; vulnery; anticonvulsant; antibacterial;		
KW	antiparasitic; cardiac; gene therapy; cancer; immune disorder;		
KW	cardiovascular disorder; neurological disease; infection; human; ss.		
XX			
OS	Homo sapiens.		
XX			
XX	WO200155308-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01309.		
XX			
RR	31-JAN-2000; 2000US-0179065.		
RR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		

CC The invention relates to human polynucleotides (AA163803-AA164012) and
CC the encoded proteins (AA434497-AA4360) useful for preventing, treating
CC or ameliorating medical conditions e.g. by protein or gene therapy. The
CC genes were isolated from a range of human tissues disclosed in the
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,
CC bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies;
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIP0 at ftp.wip0.int/pub/published_pct_sequences.

XX Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;

Query Match 56.1%; Score 2043.8; DB 22; Length 3466;

Best Local Similarity 85.2%; Pct. No. 0; Mismatches 387; Indels 18; Gaps 3;

Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

QY 1 ATGGCTCAGTTTCCACACCTTGGGTAGCTGATGTCGGGCACTTAACCTTGAG 60
DB 277 ATGGCTCAGTTTCCACACCTTGGGTAGCTGATGTCGGGCACTTAACCTTGAG 336
QY 61 GAAAGGGCCAGAGCATGACGAGCTTCTTGGCTGAGCCGATAGCGGATTTATCT 120
DB 337 GAAAGGGCCAGAGCATGACGAGCTTCTTGGCTGAGCCGATAGCGGATTTATCT 396
QY 121 GGTGATCAGCCAGGAACTTTTTCATCTGGGTACCTGAGCTGCTTGAAGCA 180
DB 397 GGTGATCAGCCAGGAACTTTTTCATCTGGGTACCTGAGCTGCTTGAAGCA 456
QY 181 ATATGGGGCTAGCGGACATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 457 ATATGGGGCTAGCGGACATGATGATGATGATGATGATGATGATGATGATGAT 516
QY 241 GGCATGAACTTATCAAACTGAAAGTACAAAGATATCAGTCCCTCCACACTTCCCT 300
DB 517 GGCATGAACTTATCAAACTGAAAGTACAAAGATATCAGTCCCTCCACACTTCCCT 576
QY 301 GTCATGAAACAGCAACAGTGTGATTTTCAAGTGCACAGCATTTGGTATGAGGAG 360
DB 577 GTCATGAAACAGCAACAGTGTGATTTTCAAGTGCACAGCATTTGGTATGAGGAG 636
QY 361 GGTAGCATGCAACATCTCAGCTGTGCTCCCTGTCGCAATGGGCTCCATTCAGTTGT 420
DB 637 GGTAGCATGCAACATCTCAGCTGTGCTCCCTGTCGCAATGGGCTCCATTCAGTTGT 696
QY 421 GGAATGTCTCAACCTTATGATCTTCTGCTCCCTCCAGCAGAGTGTCCCTGCTTAAC 480
DB 697 GGAATGTCTCAACCTTATGATCTTCTGCTCCCTCCAGCAGAGTGTCCCTGCTTAAC 756
QY 481 GGGGCTCTCCGTCAATACAGCTTGTGCTGCTTGGCGATCTTCAAGCAGACATGGCA 540
DB 757 GGGGCTCTCCGTCAATACAGCTTGTGCTGCTTGGCGATCTTCAAGCAGACATGGCA 816
QY 541 AAGAGTTCTCTTCCAGAGATGCTGTCAGGGTCACTTAAGCACTAAGTTTCAAG 600
DB 817 AAGAGTTCTCTTCCAGAGATGCTGTCAGGGTCACTTAAGCACTAAGTTTCAAG 876
QY 601 GCAATATCATTCGATGTCGACAGGCGCTCCAGCAGAGAAATGGCTGTGCTCAGTCA 660
DB 877 GCAATATCATTCGATGTCGACAGGCGCTCCAGCAGAGAAATGGCTGTGCTCAGTCA 936
QY 661 TCAAGGCTGAATACAGGAGTTTCAACAGCAGCAAACTATGAGTGAACCTTA 720
DB 937 TCAAGGCTGAATACAGGAGTTTCAACAGCAGCAAACTATGAGTGAACCTTA 996

QY 721 AC-----AGTCCCGAGGCAAGATATTCATGCAATCAAGTTTACCCAG 768
DB 997 AAGAGTTCTCTTCCAGAGATGCTGTCAGGGTCACTTAAGCACTAAGTTTCAAG 1056
QY 769 GGTGATGCTGCTTCAATATGATGATCTTCTGATGATGATGATGATGATGATGAT 828
DB 1057 GGTGATGCTGCTTCAATATGATGATCTTCTGATGATGATGATGATGATGATGAT 1116
QY 829 GAAAGATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 888
DB 1117 GAAAGATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1176
QY 889 CCCGCTCTCTCCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 948
DB 1177 CCGCTCTCTCTCCAGATATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1236
QY 949 ATGTCGTCATGATGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1008
DB 1237 ATATCTGTCATGATGATCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1296
QY 1009 GATGACGACGAC---CAGAGAGAACTGCTGTCATCTTGAATGATGATGATGATGAT 1065
DB 1297 GATGACGACGAC---CAGAGAGAACTGCTGTCATCTTGAATGATGATGATGATGAT 1356
QY 1066 AACTTCGAGCGAGCGAGTGTGAGCTGAGAGAGCGCGCGAGCGCTTGGAGCAGAG 1125
DB 1357 AACTTCGAGCGAGCGAGTGTGAGCTGAGAGAGCGCGCGAGCGCTTGGAGCAGAG 1416
QY 1126 CGCAAGAGCAGAGCGAGCTGCTGAGCGCGCGCGAGCGCGAGCGAGCGAGCGAG 1185
DB 1417 CGCAAGAGCAGAGCGAGCTGCTGAGCGCGCGAGCGCGAGCGAGCGAGCGAGCGAG 1476
QY 1186 GAGCGCGAGCAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1245
DB 1477 GAGCGCGAGCAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1536
QY 1246 CGGAGCTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1305
DB 1537 CGGAGCTGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1596
QY 1306 GCAAAACGGGAACTGAAAGGCGAGCAACTTGAATGGGAAACGAGACGAGAA 1365
DB 1597 GCAAAACGGGAACTGAAAGGCGAGCAACTTGAATGGGAAACGAGACGAGAA 1656
QY 1366 CTCCTGATCAGAGGAAACAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1425
DB 1657 CTCCTGATCAGAGGAAACAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1716
QY 1426 ACTCTGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1485
DB 1717 ACTCTGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776
QY 1486 CAGATATCAGGCTGTCAGCTGCAACCCAGAGCGAGAAATTTGAGAGCAGAACTCT 1545
DB 1777 CAGATATCAGATGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1836
QY 1546 AAGAGCTGAAGATTTGCTGAATTCACCTTACAGCAGCAGTGTGAGAACTCAGCAA 1605
DB 1837 AAGAGCTGAAGATTTGCTGAATTTCCAGAAATTCACCTTACAGCAGCAGTGTGAGAA 1896
QY 1606 ATGCTTGAAGCTTATTCAGAGAAACAGATCACTGATGATGATGATGATGATGAT 1665
DB 1897 ATGCTTGAAGCTTATTCAGAGAAACAGATCACTGATGATGATGATGATGATGAT 1956
QY 1666 CAGAACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1725
DB 1957 CAGAACTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2016
QY 1726 CTGGCCCGGAGCAGCTCGGAGAGAGCTGCAAGAGTGTGAGAGAGAGAGAGAGAG 1785
DB 2017 CTGGCTCGGAGCAGCTCGGAGAGAGAGCTGCAAGAGTGTGAGAGAGAGAGAGAG 2076
QY 1786 CTGCAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845

```

Db      2077  CTAAGGAGATTGATTTTCAATTAATCAGCTGAAGAACTAAGAAATACATATAG 2136
Qy      1846  CAGCAATCTCAAGACAGAGGTCCCTGAGGACGCCCACTGAAGCAGAAAGCAGAG 1905
Db      2137  CAACAATCTCAAGACAGAAAGCCATGAGGCTGAACCACTGAACCAAGAAACAA 2196
Qy      1906  AGGAAGACCTGAGATTAGAGCAAAAGAGCGCTCAGAGCGAGTTCAAGAAAG 1965
Db      2197  CGAAGATCATATGAAATTAAGAAAAAAGAGAGCCAAAGCGAGCTCAGAAAG 2256
Qy      1966  GACAAGCAATGAGCTGAGACATGTGACAGCAGAG--GAGACGCCAGCCCCG 2022
Db      2257  GACAAGCAATGAGCTGAGACATGTGACAGCAGAGCGAGCATCAGAGACCA 2316
Qy      2023  CACGAGAGAGCAGACTGAAGAGGAGAGACAGTGTCAAGAAAGAGAGCGAG 2082
Db      2317  CACGAAAGAGAAAACTGAAAAAGGAGAGAGTGTCAAAAAAGAGATGCGAG 2376
Qy      2083  GCCAAGCCGGAATTCAGAGACAGAGAGTGGCTTTCCATCCGATCAGAGG 2142
Db      2377  GCGAAACAGAAAGCAGAAAGACAGTGGGTGGCTTTTCATCAACACCAAG 2436
Qy      2143  AAGCTGACCAAGCAGCAGCTGTCTACCAAGAGAAAGCGCGCTTACCAT 2202
Db      2437  AAGCAGCTGTCCAGAGCAGCAGCTGTCTACCAAGAGAAAGCTTACCAT 2496
Qy      2203  CAGGAGAGTGAAGAGTGTATATTAACGAGCGCTGTACCCCTTGAATCA 2262
Db      2497  CAGGAAATGTAAAGTGTGTATTAACGAGCGCTGTACCCCTTGAATCA 2556
Qy      2263  GATGAGATCACCATCAGCAGAGAGATATAGTATGATGTGAAGAGCAGAT 2322
Db      2557  GATGAATCATATCAGCAGAGAGATATAGTATGATGTGAAGAGCAGAT 2616
Qy      2323  CCAGATGAGCTTGGAGAGAGCTGAAGAGAGCGAGTGTCTCCCTGCAAA 2382
Db      2617  CCGGCTGCTTGGAGAGAGATTAAGAAAGAGAGGTGTCTCCCTGCAAA 2676
Qy      2383  GAAAGATTCAGAAATGAGAGTTCACCTCAGCAGCAACAGATGATGATCT 2442
Db      2677  GAGAAATTCAGAAATGAGAGTTCACCTCAGTGAACAGATGATGATCAAC 2736
Qy      2443  GCGCTGCGCCCAAGCTGTGTGTGAAGCCCTGCTCCCTGAGTGAACCT 2502
Db      2737  GCGCTGCGCCCAAGCTGTGTGTGAAGCCCTGCTCCCTGAGTGAACCT 2796
Qy      2503  GAGCCTTCAGAACCCCAAGCACTGAGCAGACTTCAGTTCACAGTGGCC 2562
Db      2797  GAGCCTTCAGAACCCCAAGCACTGAGCAGACTTCAGTTCACAGTGGCC 2856
Qy      2563  AACGAGAACCAAGAAAGCAGCACTGAGATAGTGTGTGTGTGTGTGTGT 2622
Db      2857  AATGAGAAACCAAGAAAGCAGTATGAGTGTGTGTGTGTGTGTGTGTGT 2916
Qy      2623  CCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2682
Db      2917  CCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2976
Qy      2683  TCCCATCTCCGCTCTGAGGAGGATGA 2711
Db      2977  TCCCGTCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3005

```

RESULT 11
AAZ34573
ID AAZ34573 standard; cDNA; 2079 BP.

AC AAZ34573;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human SH3D1A cDNA clone 5.

```

XX      SH3D1A gene: human; Down's syndrome; leukaemia; cancer;
KW      megakaryocytic abnormality; myeloproliferative disorder;
KW      platelet disorder; neural disorder; thrombocytopenia;
KW      haematopoietic disorder; cognitive dysfunction; microcephaly;
KW      lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
XX      ss.
XX      Homo sapiens.
XX      Key Location/Qualifiers
FH      CDS 136..2079
FT      /*tag= a
XX      MO9953062-A2.
XX      PD 21-OCT-1999.
XX      PF 16-APR-1999; 99MO-US08371.
XX      ER 16-APR-1998; 98US-0082007.
XX      PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX      PI Korenberg JR, Chen X;
XX      DR WP1: 1999-633829/54.
XX      DR P-PSDB; AAY32157.
XX      PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT      the diagnosis and treatment of myeloproliferative disorders and
XX      leukaemia -
XX      Claim 2; Fig 12; 99BP; English.
XX      This is the nucleotide sequence of a cDNA clone, termed clone 5,
XX      corresponding to a novel human SH3 gene, termed the SH3D1A gene,
XX      that contributes to the development of platelets and in particular those
XX      pathogenesis of leukaemias, both in general and in particular those
XX      involving the megakaryocytic lineage. The SH3D1A gene maps to the
XX      small candidate region for low platelets on chromosome 21.
XX      Sequencing of 5 different sizes of cDNA clone from foetal brain
XX      (see AAZ34570-74) suggests that at least 3 isoforms exist. The
XX      invention provides methods for the diagnosis and treatment of
XX      megakaryocytic abnormality, myeloproliferative disorder, platelet
XX      disorder, acute leukaemia, neural disorders, thrombocytopenia,
XX      platelet disorder on chromosome 21, low platelets in deletion for
XX      21, association of gains in chromosome 21 with leukaemias, neural
XX      abnormalities, dysfunctions and disorders including brain
XX      malformations and corresponding cognitive dysfunctions,
XX      microcephaly, lissencephaly, and colpocephaly. Methods are also
XX      provided for: suppressing cells unable to regulate themselves;
XX      screening for a somatic alteration in the SH3D1A gene; monitoring
XX      the progress and adequacy of a treatment; monitoring tumour risk
XX      progress or megakaryocytic abnormality, myeloproliferative disorder,
XX      haematopoietic disorder, platelet disorder or leukaemia; and
XX      treatment of a subject (including a prenatal subject) having
XX      megakaryocytic abnormality, myeloproliferative disorder,
XX      platelet disorder, leukaemia or neural disorder using a
XX      nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX      Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;

```

Query Match 41.0%; Score 1495; DB 20; Length 2079;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

```

Qy      1  ATGGCTCAGTTTCCACACCTTTCGGTGTAGCTGTGATGTGGCGCATTAAGTGGAG 60
Db      136  ATGGCTCAGTTTCCACACCTTTCGGTGTAGCTGTGATGTGGCGCATTAAGTGGAG 195
Qy      61  GAAAGGCCAAGCATGACGAGTTCCTTAGCCTGAAGCGGATGATTTTAACT 120

```

Db 196 GAAAGACGAGCATGATCAGCAGTTCCATAGTTTAAAGCAATATCTGATTCATTACT 255
 Qy 121 GGTATCAAGCGAGAACTTTTTCATCTGGGTTACTCAGCTGCTTAGACAA 180
 Db 256 GGTATCAAGCTTGAATCTTTTTCATCTGGGTTACTCAGCTGCTTAGACAA 315
 Qy 181 ATATGGGCGCTAGCGAGCATGATTAACGATGAAAGATGATCAATGAAATTTTTCATA 240
 Db 316 ATATGGGCGCTAGCTGATGATTAATGATGAAAGATGATCAATGAAATTTTTCATA 375
 Qy 241 GCCATGAAGCTTATCAACTGAAAGTACAGATATACGCTCCCTCCACACTTCCCTC 300
 Db 376 GCTATGAACCTTATCAACTGAAAGTACAGATATACGCTCCCTCCACACTTCCCTC 435
 Qy 301 GTCATGAACAGCAACGATGGCTATTTCCAGTGCACGAGCATTTGATAGAGGAT 360
 Db 436 GTCATGAACAGCAACGATGGCTATTTCCAGTGCACGAGCATTTGATAGAGGAT 495
 Qy 361 GGTAGCATGCGACCACTCAGCTGTGCTCCTGCTCCATGGGCTCCATTCCAGTTGTT 420
 Db 496 GCGAGATGCGACCGCTTACAGCTGTGCTCCAGTGCACATGGGATCCATTCCAGTTGTT 555
 Qy 421 GGAATGTCTCCACCTTATGATCTTGTCTCCTCCAGCAGAGTGCCTCCCTGGCTAAC 480
 Db 556 GGAATGTCTCCACCTTATGATCTTGTCTCCTCCAGCAGAGTGCCTCCCTGGCTAAC 615
 Qy 481 GGGGCTCCCTCCGCTCATAGCCTTGCCTGGCTTGGCCATCTGACACCATATGGCCA 540
 Db 616 GGGGCTCCCTCCGCTTATACACTGCTGCTGCTATTTGCTATCTGACACCATATGGCCA 675
 Qy 541 AAGAGTTCTCTTCCAGCAGATCTGCTCAGGGTCACTAACTAACTAACTAACTAACTAA 600
 Db 676 AAGAGTTCTCTTCCAGCAGATCTGCTCAGGGTCACTAACTAACTAACTAACTAACTAA 735
 Qy 601 GCACAATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAGATGGCTGTGCTCAGTCA 660
 Db 736 GCACAATCATTTGATGTGCGCAGCGCTGCCACAGTGGCAGAGGGCTGTGCTCAGTCA 795
 Qy 661 TCAAGGCTGAATATCAGGAGTATTCACAGCAGCAGCAAACTATGATGACACTTA 720
 Db 796 TCAAGGCTGAATATCAGGAGTATTCACAGCAGCAGCAAACTATGATGACACTTA 855
 Qy 721 AAGAGTCCCGCAGGCAAGACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780
 Db 856 AAGAGTCCCGCAGGCAAGACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 915
 Qy 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAACTCACTGACAGAGAAATTTATC 840
 Db 916 TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAACTCACTGACAGAGAAATTTATC 975
 Qy 841 CTAGCTATGCACTTAATTTGATGTGCGATGTCTGATCAGCTGCGCCGCTCTGCT 900
 Db 976 CTAGCTATGCACTTAATTTGATGTGCGATGTCTGATCAGCTGCGCCGCTCTGCT 1035
 Qy 901 CCGAATATCATCTCTCTCTCTCAGAGATTCGCTCGGCGAGTGGATGTCCTGATATA 960
 Db 1036 CCGAATATCATCTCTCTCTCTCAGAGATTCGCTCGGCGAGTGGATGTCCTGATATA 1095
 Qy 961 AGCTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGCTGCTCAGAGATGAGCAGCAG 1020
 Db 1096 AGCTCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGCTGCTCAGAGATGAGCAGCAG 1155
 Qy 1021 C---CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGGAATTTCCAGGCA 1077
 Db 1156 CAATTATGAAGAAATTTACCTGTAAAGTTGAAGATTAAGAGCGGGAATTTGAAGCT 1215
 Qy 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTTTTGAAGCAGAGCGCAAGAGCAG 1137
 Db 1216 GGCAGTGTGAGCTGAGAGAGCGCGCAAGCGCTTTTGAAGCAGAGCGCAAGAGCAG 1275
 Qy 1138 GAGCGGTTGGCTTCAAGTGAAGCGCGCGCAGCAGAGAGAGAGAGAGCGGAGCGCCAGAG 1197
 Db 1276 GAGCGGTTGGCTTCAAGTGAAGCGCGCGCAGCAGAGAGAGAGAGAGCGGAGCGCCAGAG 1335

Qy 1198 CAGAGGCGCAAGCGGAGCTGAGAGTGGAGAGAGCAGCTGAGAGAGAGAGAGAGAGAGAG 1257
 Db 1336 CAGAGGCGCAAGCGGAGCTGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
 Qy 1258 CCGCAGCGAG 1317
 Db 1396 CCGCAGCGAG 1455
 Qy 1318 CTGGAAGGCGAG 1377
 Db 1456 CTGGAAGGCGAG 1515
 Qy 1378 AGGAG 1437
 Db 1516 AGGAG 1575
 Qy 1438 GAGTTAGAGAGCTGAG 1497
 Db 1576 GAGTTAGAGAGCTGAG 1635
 Qy 1498 TGTGCACTGGCAACCGAG 1557
 Db 1636 TGTGCACTGGCAACCGAG 1695
 Qy 1558 ATTGCTGAAGATCACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
 Db 1696 ATTGCTGAAGATCACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1755
 Qy 1618 CTTATTTCCAG 1677
 Db 1756 CTTATTTCCAG 1815
 Qy 1678 CATAGAGACTGCTTCTTACCTTCAAAAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
 Db 1816 CATAGAGACTGCTTCTTACCTTCAAAAGAGCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875
 Qy 1738 CAGCTCCGAG 1797
 Db 1876 CAGCTCCGAG 1935
 Qy 1798 GATGTTTTCACACACAGCTGAG 1857
 Db 1936 GATGTTTTCACACACAGCTGAG 1995
 Qy 1858 AAGCAGAGCTCCCTGAG 1917
 Db 1996 AAGCAGAGCTCCCTGAG 2055
 Qy 1918 GAGTTAG 1940
 Db 2056 GAGTTAG 2078
 RESULT 12
 AAK94139
 ID AAK94139 standard; cDNA, 2131 BP.
 AC AAK94139;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human full-length cDNA, SEQ ID NO: 2646.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 OS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000BP-0114089.

XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Igogai T, Hayashi K, Iehi S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR P-PSDB; AAM93229.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX Claim 8; SEQ ID NO 2646; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 other;

Query Match 40.7%; Score 1483.8; DB 22; Length 2131;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

QY 1 ATGGCTCAGTTTCCCAACCTTTGGTGTAGCTGTAGTCTGGGCGCAATACCTGTAG 60
DB 205 ATGGCTCAGTTTCCCAACCTTTGGTGTAGCTGTAGTCTGGGCGCAATACCTGTAG 264
QY 61 GAAAGGCGCAAGCATGATGACGAGTCTTCTAGCTGAGCGATGCGGGAATTATTA 120
DB 265 GAAAGGCGCAAGCATGATGACGAGTCTTCTAGCTGAGCGATGCGGGAATTATTA 324
QY 121 GGTGATCAAGCAGGAACTTTTTCATCTGCGTTACTCAGCTGCTGTACGACAA 180
DB 325 GGTGATCAAGCAGGAACTTTTTCATCTGCGTTACTCAGCTGCTGTACGACAA 384
QY 181 ATATGGGCGCTAGCGGACATGATTAATGATGAGATGAGATGAGATGAGATGAG 240
DB 385 ATATGGGCGCTAGCGGACATGATTAATGATGAGATGAGATGAGATGAGATGAG 444
QY 241 GCGATGAAGCTTATCAAACTGAAGTACAGGATATCAGCTCCCTCCCACTTCCCT 300
DB 445 GCGATGAAGCTTATCAAACTGAAGTACAGGATATCAGCTCCCTCCCACTTCCCT 504
QY 301 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACACGATTTGGTATGAGAGG 360
DB 505 GTCATGAACAGCAACAGTGGCTATTTCCAGTGCACACGATTTGGTATGAGAGG 564
QY 361 GCTACAGTCCCACTGACAGCTGTGTCTCTGCGCAATGGGCTCCATTCATTTGT 420
DB 565 GCTACAGTCCCACTGACAGCTGTGTCTCTGCGCAATGGGCTCCATTCATTTGT 624
QY 421 GGAATGCTCCCACTGATATCTTCTGCTCCTCAGCAGCAGTCTCCCTGCTGCTAC 480
DB 625 GGAATGCTCCCACTGATATCTTCTGCTCCTCAGCAGCAGTCTCCCTGCTGCTAC 684
QY 481 GGGGCTCTCCCGTATACAGCTCTGCTGCGTTGGCGATCTGACAGCCAGATGGCCA 540
DB 685 GGGGCTCTCCCGTATACAGCTCTGCTGCGTTGGCGATCTGACAGCCAGATGGCCA 744

QY 541 AAGAGTCTTCTCTCAGCAGATCTGTCCAGGGTCAACATTTAAACATTAAGTACAGAG 600
DB 745 AAGAGTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 804
QY 601 GCACATCATTCGATGTCGCGCAGGCGCTCCAGCAGCAGATGCGTGTGCTCACTCA 660
DB 805 GCACATCATTCGATGTCGCGCAGGCGCTCCAGCAGCAGATGCGTGTGCTCACTCA 864
QY 661 TCAGGCTGAATATCAGCAGTATTTCAACAGCCAGCAAACTATGATGACACTTA 720
DB 865 TCAGGCTGAATATCAGCAGTATTTCAACAGCCAGCAAACTATGATGACACTTA 924
QY 721 ACAGTCCCGCAGCAGCAATCTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780
DB 925 ACAGTCCCGCAGCAGCAATCTTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 984
QY 781 TCAATATGAAATCTTCTGACATTTGATCAAGATGAGAACTCACTGACAGAAATTTATC 840
DB 985 TCAATATGAAATCTTCTGACATTTGATCAAGATGAGAACTCACTGACAGAAATTTATC 1044
QY 841 CTAGCTATGACACCTAATTTGATGTCATGTCATGTCATGTCATGTCATGTCATGTC 900
DB 1045 CTAGCTATGACACCTAATTTGATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1104
QY 901 CCAGATATACATCCCT 960
DB 1105 CCAGATATACATCCCT 1164
QY 961 AGCTCTTCT 1020
DB 1165 AGCTCTTCT 1224
QY 1021 C---CAGAGAAAGAACTGCTCTGACATTTGAAGTAAAGAGGAGAACTTCTGAGCGA 1077
DB 1225 CAATTAAGAAAGAAATTAATCTCTGTAACGTTGAAGATTAAGAGCGGGAATTTGAACGT 1284
QY 1078 GCGAGTGTGAGCTGAGAGAGCGCGCAAGCGCTTTTGAAGCAGCAGCCCAAGAGCAG 1137
DB 1285 GCGAGTGTGAGCTGAGAGAGCGCGCAAGCGCTTTTGAAGCAGCAGCCCAAGAGCAG 1344
QY 1138 GAGCGTGTGCTCAGCTGAGAGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAG 1197
DB 1345 GAGCGTGTGCTCAGCTGAGAGCGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
QY 1198 CAGAGGCGCAAGCGGCGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
DB 1405 CAGAGGCGCAAGCGGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
QY 1258 CCGCAGCAG 1317
DB 1465 CCGCAGCAG 1524
QY 1318 CTGAAAGCGCAGCAGCAATTTGAATGGGAACGAAACCGGAGACAGAACTCTGAATCAG 1377
DB 1525 CTGAAAGCGCAGCAGCAATTTGAATGGGAACGAAACCGGAGACAGAACTCTGAATCAG 1584
QY 1378 AGGAACAG 1437
DB 1585 AGGAACAG 1644
QY 1438 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
DB 1645 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAGAGAGAGAGAGAGAGAGAG 1704
QY 1498 TGTGAGTGGCAACCCAGAGCAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
DB 1705 TGTGAGTGGCAACCCAGAGCAGAAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1764
QY 1558 ATTGCTGAATATCACTCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
DB 1765 ATTGCTGAATATCACTCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1824
QY 1618 CTTATTCAGAGAGAAACAGATCTCAGTGAACAGATTAAACAGATTCAGAGAGAGAGAG 1677

Db 1825 CTTATTCAGAAAACAGATCACTCAATGACCAATTTAAACAACTTCAGACGAAACAGTTTG 1884
 QY 1678 CATAGAGATCGCTTCTTCACTCCCTCAAAAAGAGCTTGGAGCAAAAGAGCTGCCCGGCGAG 1737
 Db 1885 CACAGAGATTCACCTTTTCACTTTAAAGAGCCCTTGGAGCAAAAGAGCACTAGCTCCGCGAG 1944
 QY 1738 CAGCTCCGGAGAGAGCTGACGAGGTGAGAGAGAGACCGAGTCAAAAGCTGCGAGAGATT 1797
 Db 1945 CACCTTCAGAGACCACTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATT 2004
 QY 1798 GATGTTTCAACCAACAGCTGAAAGAACTGAGAGAGATTCATGCAAAACAGCAACTCCAG 1857
 Db 2005 GATATTTTCAATATCACTGAGAGAGAACTAAAGAAATACACATTAAGCAACAACTCCAG 2064
 QY 1858 AAGCAAGATCTCTTGGAGGAGCGCCGATGAGACAGAGAGAGAGAGAGAGAGAGAGCTTG 1917
 Db 2065 AAGCAAAAGTCCATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATT 2124
 QY 1918 GAGTTAG 1924
 Db 2125 GAATTAG 2131
 RESULT 13
 AAS84762
 ID AAS84762 standard; cDNA; 2874 BP.
 AC AAS84762;
 DT 13-FEB-2002 (first entry)
 DE DNA encoding novel human diagnostic protein #20566.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSBQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG20575.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PS Claim 1; SEQ ID No 20566; 103bp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;
 Query Match 35.5%; Score 1291.4; DB 23; Length 2874;
 Best Local Similarity 76.0%; Pred. No. 2.1e-298;
 Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;
 QY 1303 GCCGCAAAACGGGAACTGGAAAAGCGACGACAACTTGAATGGGAAACGAGACGAGACG 1362
 Db 1 GCTGCAAAACGGGAACTGAAAAGCGACGACAACTTGAATGGGAAACGAGACGAGACG 60
 QY 1363 GAACTCTGTAATCAGAGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1422
 Db 61 GAACTCTGTAATCAGAGAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 1423 AAGACTCTGAGATTGAGTTAGTAAAGCTGTAATGACAAAAGCTAGCTAAGAGAGAAA 1482
 Db 121 AAGACTCTGAGATTGAGTTAGTAAAGCTGTAATGACAAAAGCTAGCTAAGAGAGAAA 180
 QY 1483 CTTACAGATATCAGAGTGTGAGCTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542
 Db 181 CTTACAGATATCAGAGTGTGAGCTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 QY 1543 TCTAGAGAGTGAAGATTGCTGAATACCCACTTAACAGAGAGAGAGAGAGAGAGAGAG 1602
 Db 241 TCTAGAGAGTGAAGATTGCTGAATACCCACTTAACAGAGAGAGAGAGAGAGAGAGAG 300
 QY 1603 CAAATGCTTGAAGAGTATTTCCAGAGAAACAAGTACTGAGTGAACAGTTAAACAAGTC 1662
 Db 301 CAAATGCTTGAAGAGTATTTCCAGAGAAACAAGTACTGAGTGAACAGTTAAACAAGTC 360
 QY 1663 CAGCAGAAACAGTTTGATGAGAGAGTCTGCTTCAACCTCAAAAGAGCTTGAAGAGAAAG 1722
 Db 361 CAGCAGAAACAGTTTGATGAGAGAGTCTGCTTCAACCTCAAAAGAGCTTGAAGAGAAAG 420
 QY 1723 GAGCTGGCCGCGAGAGAGTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782
 Db 421 GAGCTGGCCGCGAGAGAGTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
 QY 1783 AAGCTGCAAGAGATTTGATTTTCAACAACAGCTGAGAGAGAGAGAGAGAGATCACTAGC 1842
 Db 481 AAGCTGCAAGAGATTTGATTTTCAACAACAGCTGAGAGAGAGAGAGAGAGATCACTAGC 540
 QY 1843 AAACAGCAACTCCAGAGAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902
 Db 541 AAACAGCAACTCCAGAGAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
 QY 1903 GAGAGAGAGAGCTGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1962
 Db 601 GAGAGAGAGAGCTGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 1963 AAGGACAGAGATGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2019
 Db 661 AAGGACAGAGATGCTGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 QY 2020 CCCACAGAGAGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2079
 Db 721 CTCACAGAGAGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 QY 2080 AAGGACAGAGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2139
 Db 781 AAGGACAGAGAGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840

```

Oy 2140 GCTAAGTGGCAACCCAGGACCCCTGCTCTACCAAGAGAAAGCCCGCTTACATTTCT 2139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 GCTAAGCCAGCTGTCTCAGGCACTCTGTCTCACTGACAGAAAAAGGCTCCATTAACATTTCT 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2200 GCACAGAGAGTGTAAAGTGTATATTAACGAGCGCTGTACCCCTTTGAAATCCAGAAAT 2259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 GCACAGAGAAATGTAAAGTGTATTAACGAGCGCACTGTACCCCTTTGAAATCCAGAAAT 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2280 CACATGATATACATCTCAGCAGAGATATATGATCATGTGTGTATGAAAGCCAGCTGGA 2319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 961 CATGTGAAATCATCTATCAGCAGAGACATATGTCATGTGTGTATGAAAGCCAACTGGA 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2320 GAGCCAGATGCTGTGAGAGAGCTGAAAGGAAAGAGCGGATGTGTTCCCTGCAAACTAT 2379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1021 GAACCCGCTGCTGTGAGAGAAATTAAGAAAGAAAGAGGTTGTTCCCTGCAAACTAT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2380 GCAGAAAAGATTCCAGAAAATGAGGTTCCATCTCAGCCAAACAGTACCGATCTGACA 2439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1081 GCAGAGAAATATCCAGAAAATGAGGTTCCCGCTCCAGTGAACAGTATGATTCACACA 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2440 TCTGCCCCCTGCCCCCAACTGCTGTGCTGTGAGAACCCCTGCTCTTGTGCAATGACCTCT 2499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1141 TCTGCCCCCTGCCCCCAACTGCTGTGCTGTGAGAACCCCGCTTGTGCAATGACCTCT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2500 TCTGAGCCCTCCACAACCCCAACAACCTGAGGAGACTTCACTTCCAGTGGCCAGCAGC 2559
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1201 TCAGAGCCCTCCACGACCCCTTAATATCTGTGGCCGACTTCACTTCCAGTGGCCAGCAGC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2560 TCMAACGAGAACCCAGAAACCGACCAACTGGGATACGTGGGCGGCTCAGCTTTCTGAC 2619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1261 ACGAATGAGAAACCGAAACCGATATCTGGGATGATGGGCAAGCCCTCTCTCAC 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2620 GTACCTATGCTGTGCGCACTTACGAGAGATCAGCTTACCCCAAGCCAGCAGCCTGGC 2679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1321 GTTCCAACTGCGGCACTTAAGGAGAGTCCGCTTATCTCCAGCCAGCCCACTGGC 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2680 TCTTCCCATCTCCCGCTCTGAGGCAAGGATGAAAGGTGAAAGGCTCAAGCCGCAAGCC 2739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1381 TCTTCCCGCTCTCTGTGTAGCCAGGCTGAAAAGGTGAGAGGCTCAAGCTCAAGCC 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2740 CTGTATCCCTGAGAGCCAAAAGACAAACCACTTAATTTTAAACAAAAGTACGTCATC 2799
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 CTATATCTTGGAGAGCCAAAAGACAAACCACTTAATTTTAAACAAAAGTATGTATC 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2800 ACCCTTCTGAAACACCAAGACATGTGTGTGTGAGAAAGTTCAGAGTCAAGAGGTTGG 2859
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 ACCGTCCTGGAACACCAAGACATGTGTGTGTGAGAAAGTTC----- 1543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2860 TTCCCAAGTCTTACGTAAACTCATTTCAAGGCGCGTAAAGAAATCCACAAGCATCGAT 2919
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1544 ----- 1543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2920 ACTGGCCCTACTGAAAGTCTGTAGTCTAAAGAGAGTGGCTTCCCGCGCCGCAAGCCA 2979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1544 ----- 1543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 2980 GCCATTTCCCGAGAGATTTATTTGCCATGTACATACGAGATTTCTGAGCAAGAGAT 3039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1544 ----- 1543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3040 TTAACCTTTCAGCAAGGAGATGTATGTGTGTACCAAGAAAGATGTGACTGTGAGC 3099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1591 TTAACCTTTCAGCAAGGAGATGTATTTGTGTACCAAGAAAGATGTGACTGTGAGCA 1650
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3100 GGAACGTTGGGAGCAAGTCCGAGTCTTCCCTTAACTATGTAGAGCTTAAAGATTCA 3159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1651 GGAACGTTGGGAGCAAGTCCGAGTCTTCCCTTAACTATGTAGAGCTTAAAGATTCA 1710
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3160 GAGGCTCTGGAACCTGTGAGAAACAGGAGATTAGAAAAAAACCTGAAATTTGCCAG 3219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1711 GAGGCTCTGGAACCTGTGAGAAACAGGAGATTAGAAAAAAACCTGAAATTTGCCAG 1770
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3220 GTTATTTGCTTCTACGCTGTACTGTCTCCGAAACAATCACCCTGGCTCTGGGAGCTG 3279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1771 GTTATTTGCTTCTACGCTGTACTGTCTCCGAAACAATTAATGTCAAACTTTAAGCCCGGAAACA 3399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3280 ATTCTGATCCGAAAAAAGAACCCAGGTGATGTGGAGAGAACTGCAAGCTCGAGGG 3339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1831 ATTTTGAATCCGAAAAAAGAAC----- 1852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3340 AAAAAAGCCAGATAGGTTGTTTCAAGCAAAATTAATGTCAAACTTTAAGCCCGGAAACA 3399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1853 ----- 1852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3400 AGCAAAATCAACCCCACTGAGCTACCAAGACCGCACTGACGAGTGTGCAAGTGTG 3459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1853 ----- 1852
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3460 ATCGGATGTATCATTAACCTGACCCGCAAGATGACGATGACGATGACCTTCAAGAAAGCCAG 3519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1867 ATTTGATGTATGAGCTACACCGCCAGAAATGACATGATGAGCTGCTTCAACAGGCGCAG 1926
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3520 ATCATCAACGTCCTCAACCAAGAGAGACCCGACTGTGAGAAAGAGAAAGTCACTGAGGCAA 3579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1927 ATCATCAACGTCCTCAACCAAGAGAGACCCGACTGTGAGAAAGAGAAAGTCACTGAGGCAA 1986
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3580 GTTGGGCTCTTCCCATCCATTAATGTAAAGCTGACCAAGACATGACCCGAGCAGCAA 3639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1987 GTGGGGCTCTTCCCATCCATTAATGTAAAGCTGACCAAGACATGACCCGAGCAGCAA 2046
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy 3640 T 3640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2047 T 2047
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AAH1578
ID AAH1578 standard; cDNA; 2131 BP.
XX
AC AAH1578;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15658.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUN-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saio K, Yamamoto J;
XX
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX
PT full-length cDNAs defined in the specification, and for the detection
XX
PT and/or diagnosis of the abnormality of the proteins encoded by the
XX
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15658; 2537bp + CD ROM; English.
XX
XX
CC The present invention describes primer sets for synthesizing 5602

```

CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to a
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the protein encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH16628 and
 CC AAH16633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;

Query Match 34.6%; Score 1259.2; DB 22; Length 2131;
 Best Local Similarity 81.2%; Pred. No. 9,1e-291;
 Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

QY 1 ATGGCTCAGTTTCCACACCTTTCGGTGGTACCTGATGCTGAGCCATTAAGTGGAG 60
 Db 347 ATGGCTCAGTTTCCACACCTTTCGGTGGTACCTGATGCTGAGCCATTAAGTGGAG 406
 QY 61 GAAAGGGCCAAAGATGACAGAGTCTTACGCTGGAAGCCATAGCGGAGATTATTA 120
 Db 407 GAAAGGGCCAAAGATGACAGAGTCTTACGCTGGAAGCCATTAAGTGGAGATTATTA 466
 QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTACCTCAGCCTGCTTACAGCAA 180
 Db 467 GGTGATCAAGCAGAGAACTTTTTCATCTGAGTACCTCAGCCTGCTTACAGCAA 526
 QY 181 ATATGGGGCTTACGGAGCATGATTAACGATGGAAGAGATGATCAAGTGAATTTTCCATA 240
 Db 527 ATATGGGGCTTACGGAGCATGATTAACGATGGAAGAGATGATCAAGTGAATTTTCCATA 586
 QY 241 GCCATGAAGCTATCAAACTGAAGTACAGATATACCTCCCTCCACATTCCTCCCT 300
 Db 587 GCTATGAAGCTATCAAACTGAAGTACAGATATACCTCCCTCCACATTCCTCCCT 646
 QY 301 GTCATGAAGCAGCAACAGTGGCTATTTCCAGTGCACAGCACTTGGTATAGAGGAGATT 360
 Db 647 GTCATGAAGCAGCAACAGTGGCTATTTCCAGTGCACAGCACTTGGTATAGAGGAGATT 691
 QY 361 GCTAGCAGTGCACCACTCAGCTGTGCTCTGTGCAATGGGCTTCATTCAGATTGTT 420
 Db 692 ----- 691
 QY 421 GGAATGTCTCAACCTTAGTATCTGTGCTCCAGCAGCAGTGGCTTCCCTGGCTAAC 480
 Db 692 ----- GAGCTGTGTGCTTCCCTGGCTAAC 715
 QY 481 GGGGCTCTCCGCTCATACAGCCTGCTGCTGGTGGTGGCATCTGACAGCAGCATGGCCA 540
 Db 716 GGGGCTCTCCGCTCTATACAGCCTGCTGCTGGTGGTGGCATCTGACAGCAGCATGGCCA 775
 QY 541 AAGAGTTCTTCTTACAGAGATCTGTGCTCAGGGTCACTTAACAATAAGTTACAGAG 600
 Db 776 AAGAGTTCTTCTTACAGAGATCTGTGCTCAGGGTCACTTAACAATAAGTTACAGAG 835
 QY 601 GACACATATTTGATGTGGCAGCGCCCTCCAGCAGCAGAGATGGGCTGTGCTCAGTGA 660
 Db 836 GACACATATTTGATGTGGCAGCGCGCCCTCCAGCAGCAGAGATGGGCTGTGCTCAGTGA 895
 QY 661 TCAAGGCTGAATATACAGCAGTATATCAACAGCAGCAAAAATAATAGTGGACACTTA 720

Db 896 TCAAGGCTGAATATACAGCAGTATATCAACAGCAGCAAAAATAATAGTGGACACTTA 955
 QY 721 AAGAGTTCTTCTTACAGAGATCTGTGCTCAGGGTCACTTAACAATAAGTTACAGAG 780
 Db 956 AAGAGTTCTTCTTACAGAGATCTGTGCTCAGGGTCACTTAACAATAAGTTACAGAG 1015
 QY 781 TCAATATGAATCTTCTGACATATGATCAAGTGAAGAACTCAGCAGAGAAATTTATC 840
 Db 1016 TCAATATGAATCTTCTGACATATGATCAAGTGAAGAACTCAGCAGAGAAATTTATC 1075
 QY 841 CTAGCTATGACCTTAATGATGTTGCTCATGTCAGCAGCAGCCTGCTGCT 900
 Db 1076 CTAGCTATGACCTTAATGATGTTGCTCATGTCAGCAGCAGCCTGCTGCT 1135
 QY 901 CCAAGATATCATCTCTCTTCTTCAAGAGTGTGCTCCGCAAGTGGATGCTGCTCAT 960
 Db 1136 CCAAGATATCATCTCTCTTCTTCAAGAGTGTGCTCCGCAAGTGGATGCTGCTCAT 1195
 QY 961 AGCTCTTCTGCTGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGATGAGACAG 1020
 Db 1196 AGCTCTTCTGCTGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGATGAGACAG 1255
 QY 1021 C---CAGAGAGAAATCTGCTGTGACATTTGAGATTAAGAAAGCGGAGACTTCAGAGCA 1077
 Db 1256 CAATTATGAAGAAATTTACCTGTATACCTTTGAGATTAAGAAAGCGGAGACTTCAGAGCA 1315
 QY 1078 GGCAGTGTGAGCTGTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1137
 Db 1316 GGCAGTGTGAGCTGTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1375
 QY 1138 GAGCGGTTGGCTCAGCTGTGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1197
 Db 1376 GAGCGGTTGGCTCAGCTGTGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1435
 QY 1198 CAGAGGCGCAAGCGGCACTGTGAGAGCGCTGTGAGAGCGGAGAGCTGTGAG 1257
 Db 1436 CAGAGGCGCAAGCGGCACTGTGAGAGCGCTGTGAGAGCGGAGAGCTGTGAG 1495
 QY 1258 CGGCAAGCAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGCAAGCGGCAAGCGGCA 1317
 Db 1496 CAGAGGCGCAAGCGGCACTGTGAGAGCGCTGTGAGAGCGGAGAGCTGTGAG 1555
 QY 1318 CTGGAAGGCGGCAAGCTTGAATGGGAACGGAACCGGAGACAGAACTCTGATACAG 1377
 Db 1556 CTGGAAGGCGGCAAGCTTGAATGGGAACGGAACCGGAGACAGAACTCTGATACAG 1615
 QY 1378 AGGAACAAGAGCAG 1437
 Db 1616 AGGAACAAGAGCAG 1675
 QY 1438 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAACTTCAGAGATACAG 1497
 Db 1676 GAGTTAGAGCTCTGAATGACAAAGAGATCAGCTAGAGAGAACTTCAGAGATACAG 1735
 QY 1498 TGTGCACTGGCAAGCCAGAGCAGAGAAATTTGAGAGCAGAACTGTAGAGAGCTTACA 1557
 Db 1736 TGTGCACTGGCAAGCCAGAGCAGAGAAATTTGAGAGCAGAACTGTAGAGAGCTTACA 1795
 QY 1558 ATTGCTGAAGTCAACCACTTACAGCAGAGAGTGGCAGAGATCTCAGCAATGCTTGAAGA 1617
 Db 1796 ATTGCTGAAGTCAACCACTTACAGCAGAGAGATGGTGAAGATCTCAGCAATGCTTGAAGA 1855
 QY 1618 CTTATTCAGAGAAACAGATACCTAGTACAGCAAGTTAAACAAGTCCAGCAGAAAGTTTG 1677
 Db 1856 CTTATTCAGAGAAACAGATACCTAGTACAGCAAGTTAAACAAGTCCAGCAGAAAGTTTG 1915
 QY 1678 CATAGAGCTGCTCTTACCTTCAACCTCAAGAGCTTGGAGAGAAAGAGAGCTGCGCGAG 1737
 Db 1916 CATAGAGCTGCTCTTACCTTCAACCTCAAGAGCTTGGAGAGAAAGAGAGCTGCGCGAG 1975
 QY 1738 CAGCTCCGAGAGAGAGCTGAT 1797

Db 1976 CACCTACAGACCAACTGATGAACTGAGAAAGAACTAGATCAAACTACAGAGATT 2035
 Qy 1798 GATGTTTCAACAACCCAGCTGAAGAACTGAGAGATACATAGCAAAACAGCACTCCAG 1857
 Db 2036 GATTTTTCATTAATCAGCTGAGAACTGAGAAATACCAATTAAGCAAACTCCAG 2095
 Qy 1858 AAGCAGAGTCCCTGAGAGCAGCGGACTGAAGCAG 1893
 Db 2096 AAGCAAAAGTCCATGAGAGCTGAACGACTGAAGCAG 2131

RESULT 15
 AA234574
 ID AA234574 standard; cDNA; 3231 BP.
 XX AA234574;
 AC
 XX 01-FEB-2000 (first entry)
 DT
 XX Human SH3D1A cDNA clone 9.
 DE
 XX SH3D1A gene; human; Down's syndrome; leukaemia; cancer;
 KM megakaryocytic abnormality; myeloproliferative disorder;
 KM platelet disorder; neural disorder; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
 KM ss.
 XX
 OS Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT CDS 2..1927
 FT /+tag= a
 XX MO9953062-A2.
 XX 21-OCT-1999.
 XX 16-APR-1999; 99MO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SINAI HEAL/TH SYSTEM.
 XX PA
 XX (CEDA-) CEDARS SINAI HEAL/TH SYSTEM.
 XX PI
 XX Korenberg JR, Chen X;
 DR WPI; 1999-633829/54.
 DR P-PSDB; AAY32158.
 PT Nucleic acid from the human SH3D1A gene and its products; useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia -
 XX
 PS Claim 2; Fig 14; 99p; English.

This is the nucleotide sequence of a non-full-length cDNA (clone 9) corresponding to a novel human SH3 gene, termed the SH3D1A gene, that contributes to the development of platelets and the pathogenesis of leukaemias, both in general and in particular those involving the megakaryocytic lineage. The SH3D1A gene maps to the small candidate region for low platelets on chromosome 21. Sequencing of 5 different sizes of cDNA clone from foetal brain (see AA234570-74) suggests that at least 3 isoforms exist. The invention provides methods for the diagnosis and treatment of megakaryocytic abnormality, myeloproliferative disorder, platelet disorder, acute leukaemia, neural disorders, thrombocytopenia, platelet disorder on chromosome 21, low platelets in deletion for 21, association of gains in chromosome 21 with leukaemias, neural abnormalities, dysfunctions and disorders including brain malformations and corresponding cognitive dysfunctions, microcephaly, lissencephaly, and colpocephaly. Methods are also provided for: suppressing cells unable to regulate themselves; screening for a somatic alteration in the SH3D1A gene; monitoring

CC the progress and adequacy of a treatment; monitoring tumour risk
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC treatment of a subject (including a prenatal subject) having
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC platelet disorder, leukaemia or neural disorder using a
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.
 XX
 SO Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;

Query Match 34.4%; Score 1252.8; DB 20; Length 3231;
 Best Local Similarity 77.2%; Pred. No. 37e-289;
 Matches 1652; Conservative 0; Mismatches 272; Indels 216; Gaps 2;

Qy 1506 GGCACCCAGAGGCAAGAAATGAGACGACGAACTAGAGAGCTGAAGATTGCTGA 1565
 Db 1 GACCAACCAAGGCAAGAAATGAGACGACGAACTAGAGAGCTGAAGATTGCTGA 60
 Qy 1566 AATCACCCTTACAGAGCAGGTTGAGAACTCAGCAAAATGCTTGAAGACTTATTC 1625
 Db 61 AATCACCCTTACAGAGCAGGTTGAGAACTCAGCAAAATGCTTGAAGACTTATTC 120
 Qy 1626 AGAGAAACAGATCTCAGTGAACGATTAAACAGTCCAGCAGAACTTGCATAGGA 1685
 Db 121 AGAAAAAGATCTCAGTGAACGATTAAACAGTCCAGCAGAACTTGCATAGGA 180
 Qy 1686 CTGCTCTTACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGGAGCAGACTCG 1745
 Db 181 TTCACTGTTCACCTTAAAGAGCTTGAAGCAAAAGAGCTGCGGAGCAGACTCG 240
 Qy 1746 GAGACAGCTGAGCAGGTTGAGAGAGACAGGTCAGAACTGAGAGATGATGTTTT 1805
 Db 241 AGACCAACTGGATTAAGTGAAGAAAGAACTAGATCAAACTACAGAGATGATTTTT 300
 Qy 1806 CAACACAGCTGAAGAACTGAGAGATCATAGCAAAACAGCACTCCAGAAACAGAG 1865
 Db 301 CAATATACAGCTGAAGAACTGAAGAAATACATATAGCAACATCCAGAAACAAA 360
 Qy 1866 GTCCCTGAGAGCAGGAGCTGAAGCAAAAGAGCAGAGAGAGAGCTGAGTTAGA 1925
 Db 361 GTCCATGAGAGCTGAAGCACTGAAGCAAAAGAGAGAGAGAGATCATAGATTTGA 420
 Qy 1926 GAAACAAAGAGAGAGCCTCAGAGCAGATTCAAGAAAGGAGCAAGATGCTGAGCA 1985
 Db 421 AAAACAAAGAGAGAGAGCCTCAGAGCAGATTCAAGAAAGGAGCAAGATGCTGAGCA 480
 Qy 1986 TGTGACAGAGAG--GAGCAGCAGCGCCCGGAAACCCACAGAGAGCAGACTGAA 2042
 Db 481 TGTGACAGAGAGAGCAGCAGATCAGAGCCAGAAACCTCCACAGAGAGAGAACTGAA 540
 Qy 2043 GAGGGAAGACAGTCTCAGAAAGAGAGGAGGAGAGAGAGCCAGGCGGAAATGCAAGA 2102
 Db 541 AAGGAGAGAGTCTCAAAAGAGAGTGGCAGAGAAAGGCAACAGAGAGCAAGAGA 600
 Qy 2103 CAAGCAGAGTGGCTTTTCCATCCGATCAGAGCAGCTTAAGCTGCGCACCCAGCAGC 2162
 Db 601 CAAGCTGGTGGCTTTTCCATCAACACAGAGCAGCTTAAGCTGCGCACCCAGCAGC 660
 Qy 2163 CTGCTTACACAGAGAAAGGCGCTTACATTTGTCACAGAGAGTGTAAAGTGT 2222
 Db 661 CTGCTTACACAGAGAAAGGCTTACATTTGTCACAGAGAGATGTAAAGTGT 720
 Qy 2223 AATTACCGAGCGGTGATACCCCTTGAATCCAGAAAGTCAGATGATCCATCCAGCC 2282
 Db 721 GTATTACCGGACCTGATACCCCTTGAATCCAGAAAGTCAGATGATCCATCCAGCC 780
 Qy 2283 AGAGATATAGTCAATGATGATGAAGACCAAGCTGAGAGCAGAGATGCTTGAAGAGA 2342
 Db 781 AGAGATATAGTCAATGATGATGAAGACCAAGCTGAGAGACCGGCTGCTTGAAGAGA 840
 Qy 2343 GCTGAAGAGAGAGAGGAGTGTCTCCGCAACTATGACGAGAAAGATTCCAGAAATGA 2402
 Db 841 ATTAAAGAGAGAGAGGAGTGTCTCCGCAACTATGACGAGAAAGATTCCAGAAATGA 900

```

QY 2403 GGTTCACATCCAGCCAAACAGTAGCCGATCTGACATCTGCCCCCTGCCCCAAACTGAC 2462
Db 901 GGTTCGCCGCTCAGTGAAGAACAGTAGCTGATTCACATCTGCCCCCTGCCCCAAACTGAC 960
QY 2463 TCTGCGTAGAGACCCCTGCTCTCTTGTGCGAGTAGCCTCTTGTGAGCCCTCCACAACCCCA 2522
Db 961 CTTCGGGTGAGACCCCGCCCTTTGTGCACTTAACCTCTTCAGAGCCCTCCACGACCCCTTA 1020
QY 2523 CAATCGGGGAGACTTCAGTTCCACGTCGGCCAGCACTCAAAAGAGAAAGCCAGAAACGGA 2582
Db 1021 TTAATCGGGGAGACTTCAGTTCCACGTCGGCCAGCACTCAAAAGAGAAACGGA 1080
QY 2583 CAATCGGGGAGACTTCAGTTCCACGTCGGCCAGCACTCAAAAGAGAAAGCCAGTTAG 2642
Db 1081 TTAATCGGGGAGACTTCAGTTCCACGTCGGCCAGCACTCAAAAGAGAAAGCCAGTTAG 1140
QY 2643 GCAGAGATCAGCCTTTTACCCCAAGCCACAGCCACTGCTCTCCCATCTCCGCTCTGGG 2702
Db 1141 GCAGAGATCAGCCTTTTACCCCAAGCCACAGCCACTGCTCTCCCATCTCCGCTCTGGG 1200
QY 2703 CCAGGGTGAAGAGTGAAGGGGTAGAACGCAAGCCCTGTATCCTTGAAGCCCAAAA 2762
Db 1201 CCAGGGTGAAGAGTGAAGGGGTAGAACGCAAGCCCTGTATCCTTGAAGCCCAAAA 1260
QY 2763 AGACAACCACTTAATTTTAAACAAAAGTAGCGTATCACCGTTCTGGAACAGCAAGACAT 2822
Db 1261 AGACAACCACTTAATTTTAAACAAAAGTAGCGTATCACCGTTCTGGAACAGCAAGACAT 1320
QY 2823 GTGGTGGTTTGAGAAATCAAGGTCAAGAGGTGGTTCCCAAGTCTTAAGTGAACAT 2882
Db 1321 GTGGTGGTTTGAGAAATCAAGGTCAAGAGGTGGTTCCCAAGTCTTAAGTGAACAT 1380
QY 2883 CATTTCAAGGCCCCGTGAAGAAATCCACAAGCATCGATCTGGCCCTACTGAAGTCTGC 2942
Db 1381 CATTTCAAGGCCCCGTGAAGAAATCCACAAGCATCGATCTGGCCCTACTGAAGTCTGC 1440
QY 2943 TAGTCTAAAGAGATGGCTTCCCGGCGCCCAAGCCAGCCATTCGCGGAGAAAGTTAT 3002
Db 1441 TAGTCTAAAGAGATGGCTTCCCGGCGCCCAAGCCAGCCATTCGCGGAGAAAGTTAT 1492
QY 3003 TGCCATGTACACATACGAGAGTCTGAGCAAGAGATTTAACCTTTCAGCAAGGGGATGT 3062
Db 1493 TGCCATGTACACATACGAGAGTCTGAGCAAGAGATTTAACCTTTCAGCAAGGGGATGT 1492
QY 3063 GATTGTGTTACCAAGAAAGATGTGACTGTGACGGGAAGGTGGGCGACAAGTCCGG 3122
Db 1493 GATTGTGTTACCAAGAAAGATGTGACTGTGACGGGAAGGTGGGCGACAAGTCCGG 1492
QY 3123 AGTCTTCCCTCTAATCTATGTAGAGCTTAAGATTCAGAGGGCTTGAACTGTGGGA 3182
Db 1493 AGTCTTCCCTCTAATCTATGTAGAGCTTAAGATTCAGAGGGCTTGAACTGTGGGA 1492
QY 3183 AACAGGAGTTAGGAAAAAAACCTGAATTTGCCAGTTATGTCTCTACGCTGCTAC 3242
Db 1493 AACAGGAGTTAGGAAAAAAACCTGAATTTGCCAGTTATGTCTCTACGCTGCTAC 1527
QY 3243 TGTCTCCGACAACTGACCTGCTCTTGGGCAAGTGTATCTGATCCGAAAAAGAACCC 3302
Db 1528 TGTCTCCGACAACTGACCTGCTCTTGGGCAAGTGTATCTGATCCGAAAAAGAACCC 1587
QY 3303 AGGTGGATGGTGGGAAGAGAACTGCAAGCTGAGGGAAAAAGCCAGATAGGGTGT 3362
Db 1588 AGGTGGATGGTGGGAAGAGAACTGCAAGCTGAGGGAAAAAGCCAGATAGGGTGT 1647
QY 3363 TCCAGCAATTAATGTCAAACTTCAAGCCCGGAAACAGCAAAATCACCCCACTGAGCT 3422
Db 1648 TCCAGCAATTAATGTCAAACTTCAAGCCCGGAAACAGCAAAATCACCCCACTGAGCT 1707
QY 3423 ACCCAAGACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 3482
Db 1708 ACCCAAGACCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1767

```

```

QY 3483 CCAGAACGATGACGAAGTACCTTCAGCAAAAGCCAGATCATCAAGTCTCAACAGGA 3542
Db 1768 GAGAAATGACGATGAGCTGCGCTTCAACAGGCGCAGATCATCAAGTCTCAACAGGA 1827
QY 3543 GAGCCCGACCTGGTGAAGAGAGAAATGCAAGTGGGCTCTTCCATCAATTA 3602
Db 1828 GAGCCCGACCTGGTGAAGAGAGAAATGCAAGTGGGCTCTTCCATCAATTA 1887
QY 3603 TGTAAAGCTGACACAGACATGAGACCCCAAGCCAGCAATGA 3642
Db 1888 TGTAAAGCTGACACAGACATGAGACCCCAAGCCAGCAATGA 1927

```

Search completed: December 4, 2003, 15:50:26
 Job time : 608.079 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:24:15 ; Search time 4655.38 Seconds
(without alignments)
19013.884 Million cell updates/sec

Title: US-09-674-237a-2

Perfect score: 3642

Sequence: 1 atggtcagcttccaccaccc.....tgaccaccagcagcaatga 3642

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gse_hum:*
18: em_gse_inv:*
19: em_gse_pin:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rtd:*
26: em_gse_phg:*
27: em_gse_vrt1:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	41.1	2079	11	BC013578 Homo sapi
2	1495.	41.0	2126	11	BC020269 Homo sapi
3	809.8	22.2	818	14	CA324450 UI-M-FY0-
4	795.6	21.8	911	13	BQ942708 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
5	748.8	20.6	2589	11	AK082606
6	727.2	20.0	732	14	CA750495
7	724	19.9	725	14	CB248849
8	722.8	19.8	732	12	BM950476
9	721.2	19.4	738	12	BM944544
10	704.8	19.4	729	13	BU704308
11	701.8	19.3	706	14	CB521237
12	699.4	19.2	703	14	CB527154
13	691.2	19.0	713	14	CA329050
14	684.2	18.8	688	14	CB244555
15	658.2	18.1	979	13	BQ719508
16	656.6	18.0	690	13	BQ179493
17	654.4	18.0	687	14	CB521525
18	635	17.4	636	14	CD347867
19	633.6	17.4	650	10	BB656585
20	625	17.2	636	14	CB526899
21	623	17.1	782	13	BU614534
22	609.6	16.7	616	29	CC325581
23	606.8	16.7	610	13	BQ443095
24	602	16.5	603	9	AI594919
25	598	16.4	598	14	CA873077
26	593.2	16.3	606	10	AM910367
27	592.6	16.3	957	14	CB587690
28	583.8	16.0	587	14	CB724039
29	581.8	16.0	585	14	CB518599
30	580.4	15.9	582	29	CC183329
31	577.4	15.9	579	14	CA879211
32	572	15.7	572	14	CA895623
33	571.4	15.7	573	14	CA874698
34	565	15.5	750	14	CD356862
35	563.4	15.5	773	12	BM681943
36	558.8	15.3	639	13	BO201258
37	555	15.2	1062	13	BO898718
38	550.2	15.1	555	14	CA563686
39	550	15.1	751	14	CA448081
40	547	15.0	906	13	BU519029
41	545.4	15.0	547	14	CA565513
42	539.6	14.8	608	14	CA511580
43	537.8	14.8	724	10	BR373101
44	534.8	14.7	722	14	CB521233
45	534.8	14.7	728	14	BY731854

ALIGNMENTS

RESULT 1
LOCUS BC013578
DEFINITION Homo sapiens, similar to interectin 1 (SH3 domain protein), clone IMAGE:3878242, mRNA.
ACCESSION BC013578
VERSION BC013578.1 GI:15488896
KEYWORDS
SOURCE HTC.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 2079)
Strausberg, R.
TITLE Direct Substitution
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK
COMMENTS NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILMIL ac: <http://image.jlml.gov>
 Series: IRAX Plate: 14 Row: C Column: 5
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796
 This clone has the following problem: retained intron.

FEATURES

source

1..2079

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3878242"

/issue_type="Lung, large cell carcinoma"

/clone_id="N1H MGC_68"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

BASE COUNT 703 a 451 c 492 g 433 t

ORIGIN

Query Match 41.1%; Score 1496.6; DB 11; Length 2079;
 Best Local Similarity 86.0%; Pred. No. 0;
 Matches 1671; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

QY 1 ATGGCTCATGTTCCACACCTTTGGGTGATGCTGTGAGCCCTAATCTGTGAG 60
 DB 132 ATGGCTCATGTTCCACACCTTTGGGTGAGCCCTGTGATCTGTGAGCCCTAATCTGTGAG 191
 QY 61 GAAAGGCGCAGATGACACCAAGTCTTACCTGAGCCGATGAGCCGATTTATTA 120
 DB 192 GAAAGGCGCAGATGACACCAAGTCTTACCTGAGCCGATGAGCCGATTTATTA 251
 QY 121 GGTATCAAGGAGGAACTTTTTCATCTGGGTGATCTGAGCCCTGTTAGACAA 180
 DB 252 GGTATCAAGGAGGAACTTTTTCATCTGGGTGATCTGAGCCCTGTTAGACAA 311
 QY 181 ATATGGGCGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCCATA 240
 DB 312 ATATGGGCGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCCATA 371
 QY 241 GCCATGAAGCTTATCAACTGAAGTAAAGATATGCTCCCTCCATCTTCCCT 300
 DB 372 GCTATGAAGCTTATCAACTGAAGTAAAGATATGCTCCCTCCATCTTCCCT 431
 QY 301 GTCATGAAGACCAACAGTGGCTAATTTCCAGTGACCAAGATTTGTAAGAGGAT 360
 DB 432 GTCATGAAGACCAACAGTGGCTAATTTCCAGTGACCAAGATTTGTAAGAGGAT 491
 QY 361 GCTAGCATGCCACCACTGACAGCTGTGCTCTGTGCAATGGGCTCCATTCAGATTGT 420
 DB 492 GCGACGATGCCACCACTGACAGCTGTGCTCTGTGCAATGGGCTCCATTCAGATTGT 551
 QY 421 GGAATGCTCCCAACCTGATATCTTCTGTGCTCCACAGCAGCTGTGCCCCCTGCTAAC 611
 DB 552 GGAATGCTCCCAACCTGATATCTTCTGTGCTCCACAGCAGCTGTGCCCCCTGCTAAC 611
 QY 481 GGGGCTCTCCGCTATGACAGCTGTGCTGCTGTTGGGATTCCTGACGACATGGCCA 540
 DB 612 GGGGCTCTCCGCTATGACAGCTGTGCTGCTGTTGGGATTCCTGACGACATGGCCA 671
 QY 541 AAGAGTCTTCTTACAGCAGATCTGTGCTCAGAGTCACAATTAACTAATAGTTACAGAG 600
 DB 672 AAGAGTCTTCTTACAGCAGATCTGTGCTCAGAGTCACAACTAATAGTTACAGAG 731
 QY 601 GCACATATTCATGATGTGCGCAGGCGCTCCAGACGAGATGGGCTGTGCTCAGTCA 660
 DB 732 GCACATATTCATGATGTGCGCAGGCGCTCCAGACGAGATGGGCTGTGCTCAGTCA 791
 QY 661 TCAGGCTGAATAACAGGCACTTATTCACAGCCACCAAACTATGAGTGAACACTTA 720

DB 792 TCAGAGCTGAATAACAGGCACTTATTCATGATGATACCAAACTATGAGTGAACACTTA 851
 QY 721 ACAGGCTCCCAAGCAGAACTATTTCTATGATCAATCAAGTTTACCCAGGCTCAGTGGCT 780
 DB 852 ACAGGCTCCCAAGCAGAACTATTTCTATGATGATGATGATGATGATGATGATGATGAT 911
 QY 781 TCATATGAACTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 912 TCATATGAACTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 971
 QY 841 CTACGATGACCTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 972 CTGCGAATGACCTCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1031
 QY 901 CCAGAAATACATCCCT 960
 DB 1032 CCAGAAATACATCCCT 1091
 QY 961 AGCT 1020
 DB 1092 AGCTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1151
 QY 1021 C---CAGAGAAAGAACTGCTGTCATTTGAAGATAAGAACCGGAGAACTTTCAGACGA 1077
 DB 1152 CAATTAGAAAAGAAATTAATCTGTAAGCTTTGAAGATAAGAACCGGAGAACTTTCAGACGT 1211
 QY 1078 GGCAGTGTGAGCTGAGAAACCGCCGAAAGCGCTCTTGAAGACGACCGCAAGACGAG 1137
 DB 1212 GGCAGTGTGAGCTGAGAAACCGCCGAAAGCGCTCTTGAAGACGACCGCAAGACGAG 1271
 QY 1138 GAGCGTGTGAGCTGAGAAACCGCCGAAAGCGCTCTTGAAGACGACCGCAAGACGAG 1197
 DB 1272 GAGCGTGTGAGCTGAGAAACCGCCGAAAGCGCTCTTGAAGACGACCGCAAGACGAG 1331
 QY 1198 CAGGAGGCAAGCGGAGCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAG 1257
 DB 1332 CAAGAGGCAAGCGGAGCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGAG 1391
 QY 1258 GCGCAGCGAAG 1317
 DB 1392 GCGCAGCGAAG 1451
 QY 1318 CTGGAAGGAG 1377
 DB 1452 CTGGAAGGAG 1511
 QY 1378 AGGAACAG 1437
 DB 1512 AGGAACAG 1571
 QY 1438 GAGTTAGAGCTCTGAATGACAAAAGACATGCTAGAGAGAAAATTCCAGATATCAGG 1497
 DB 1572 GAAATTAAGAGCTCTGAATGACAAAAGACATGCTAGAGAGAAAATTCCAGATATCAGG 1631
 QY 1498 TGTGCACTGCAACCCAG 1557
 DB 1632 TGTGCACTGCAACCCAG 1691
 QY 1558 ATTGCTGAATTCACCCCTTACAGCAGCAGTGTGAGAGAACTCAGCAATGCTTGAAGA 1617
 DB 1692 ATTGCTGAATTCACCCCTTACAGCAGCAGTGTGAGAGAACTCAGCAATGCTTGAAGA 1751
 QY 1618 CTATTCAGAGAAACAGATATCTGATGACCAAGTTAAACAAGTCCAGCAGAAACAGTTTG 1677
 DB 1752 CTATTCAGAGAAACAGATATCTGATGACCAAGTTAAACAAGTCCAGCAGAAACAGTTTG 1811
 QY 1678 CATAGAGCTCGCTTCTTACCTTCAAAAGAGCTTGAAGCAAGAGAGCTGCGCCGAG 1737
 DB 1812 CACAGAGATTCATCTGTTAATCTTAAAGAGCTTGAAGCAAGAGAGCTGCGCCGAG 1871
 QY 1738 CAGCTCGGAGAGAGCTGAG 1797

Db 1872 CACCTACGAGACCACTGATGAGTGGAGAAAGAACTAGATCAAACTACAGAGATT 1931

QY 1798 GATGTTTCAACACAGCAGTGAAGAACTGAGAGATCATATGCAAAAGCAACTCCAG 1857

Db 1932 GATATTTTCAATATATAGCTGAGAGAACTAGAGAAATTCACATATAGCAACTCCAG 1991

QY 1858 AAGCAGAGGTCCCTGAGGCGAGCGAGCTGAAAGCAGAAAGAGAGAGAGAGAGCTG 1917

Db 1992 AAGCAAAAGTCCATGAGGCTGAACGACTGAAACAGAAAGAAAGAGAGAGATCATAT 2051

QY 1918 GAGTTAGAGAGCAAAAGAGAA 1940

Db 2052 GAATTAGAAAAACAAAAAAA 2074

RESULT 2

LOCUS BC020269 2126 bp mRNA linear HTC 19-DEC-2001

DEFINITION Homo sapiens, clone IMAGE:4899011, mRNA.

ACCESSION BC020269

VERSION BC020269.1 GI:17939664

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2126)

AUTHORS Strausberg, R.

JOURNAL Direct Submission

Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

info@cgsc.bc.ca

Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie McPherson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>

Series: IRAL Plate: 40 Row: n Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796

This clone has the following problem: no cloning site / microdeletion.

FEATURES

source

1..2126

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4899011"

/issue_type="Pancreas", epithelioid carcinoma"

/clone_lib="NIH-MGC 42"

/lab_host="DH10B-R"

/note="Vector: POTB7"

BASE COUNT 712 a 462 c 517 g 435 t

ORIGIN

Query Match 41.0%; Score 1495; DB 11; Length 2126;

Best Local Similarity 85.9%; Pred. No. 0;

Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

QY 1 AAGGCTCAGTTTCCACACCTTTCGGTGTACCTGAGATGTCGGGCACTAGTGGAG 60

Db 182 ATGGCTCAGTTTCCACACCTTTCGGTGTACCTGAGATGTCGGGCACTAGTGGAG 241

QY 61 GAAAGGGCAACATAGACAGAGAGTTCCTTACCTGAAAGCCATAGCGGATTTTACT 120

Db 242 GAAAGGCAACATAGATAGAGAGTTCCTTACCTTAAAGCAATTCGTGATTCATCT 301

QY 121 GGTATCAAGCAGAGACTTTTTCCTCAATCTGGGTTACTGACCTGCTTACAGCA 180

Db 302 GGTATCAAGCTAGAGACTTTTTCCTCAATCTGGGTTACTGACCTGCTTACAGCA 361

QY 181 ATATGGGCGCTGCGGACATGATATGATGGAAGATGATCAATGATTTTCCATA 240

Db 362 ATATGGGCGCTGCGGACATGATATGATGGAAGATGATCAATGATTTTCCATA 421

QY 241 GCATGAGCTTATCAAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTCCCT 300

Db 422 GCTATGAACTTATCAAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTCCCT 481

QY 301 GTCATGAAACAGCAACAGTGGCTATTTTCAGTGCACACGACTTTGATAGAGAGATT 360

Db 482 GTCATGAAACAGCAACAGTGGCTATTTTCAGTGCACACGACTTTGATAGAGAGATT 541

QY 361 GCTAGCAGCCACCACTACAGCTGTGTCTCTGTGCCATAGGCTCCATTCACTTGT 420

Db 542 GCCACAGCCACCACTACAGCTGTGTCTCTGTGCCATAGGCTCCATTCACTTGT 601

QY 421 GGAATGTCACACCTTATCTTCTGTCCTCCAGAGAGAGTCCCTCCCTGCTTAC 480

Db 602 GGAATGTCACACCTTATCTTCTGTCCTCCAGAGAGAGTCCCTCCCTGCTTAC 661

QY 481 GGGGCTCTCCGCTATACAGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

Db 662 GGGGCTCTCCGCTATACAGCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721

QY 541 AAGATGTTCTCTTACAGAGATCTGTGTGAGGCTGACAACTTAACTATAGTACAG 600

Db 722 AAGATGTTCTCTTACAGAGATCTGTGTGAGGCTGACAACTTAACTATAGTACAG 781

QY 601 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGAGAGAGATGAGGCTGCTGCTGCT 660

Db 782 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGAGAGAGATGAGGCTGCTGCTGCT 841

QY 661 TCAGAGCTGAATACAGGAGATTTTCAACAGCCACGACAAACTATAGTGGACATTA 720

Db 842 TCAGAGCTGAATACAGGAGATTTTCAACAGCCACGACAAACTATAGTGGACATTA 901

QY 721 ACAAGTCCCGAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 780

Db 902 ACAAGTCCCGAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGTGGCT 961

QY 781 TCATATGGAATCTTTTTCAGATGATCAAGATGAAACTCAGCAGAGAAATTTATC 840

Db 962 TCATATGGAATCTTTTTCAGATGATCAAGATGAAACTCAGCAGAGAAATTTATC 1021

QY 841 CTAGCTATGACCTATGATGATGTCATGTCGTCAGCCAGCTCCGCTGCTGCT 900

Db 1022 CTAGCTATGACCTATGATGATGTCATGTCGTCAGCCAGCTCCGCTGCTGCT 1081

QY 901 CCAGAAATACATCCCTCTCTCTTCAAGAGAGTTGCTCCGAGAGTGGATGTCCTCATTA 960

Db 1082 CCAGAAATACATCCCTCTCTCTTCAAGAGAGTTGCTCCGAGAGTGGATGTCCTCATTA 1141

QY 961 AGCTCTTCTCTGATGATGAGAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1020

Db 1142 AGCTCTTCTCTGATGATGAGAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1201

QY 1021 C---CAGAGAGAACTGCTGTGACATTTGAAGTAAAGAGCGGAGAACTTCAGAGCA 1077

Db 1202 CAATTAGAAAGAAATTAACCTGTAAAGTTGAAGTAAAGAGCGGAGAACTTTAAAGCT 1261

```

QY 1078 GGCACTGTGAGCTGGAGAGCGCCGCAAGCGCTCTTGGACAGCAGCCGCAAGAGCAG 1137
Db 1262 GGCACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
QY 1138 GAGCGGTTGGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197
Db 1322 GAGCGGTTGGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1381
QY 1198 CAGAGAGCGCAGCGCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257
Db 1382 CAGAGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1441
QY 1258 CGGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1317
Db 1442 CGGAGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1501
QY 1318 CTGAGAGAGCAGCAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1377
Db 1502 CTGAGAGAGCAGCAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1561
QY 1378 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
Db 1562 AGGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1621
QY 1438 GAGTTAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497
Db 1622 GAGTTAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1681
QY 1498 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557
Db 1682 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741
QY 1558 ATTGCTGAATGACCCAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617
Db 1742 ATTGCTGAATGACCCAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
QY 1618 CTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 1802 CTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1861
QY 1678 CATAGAGAGCTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737
Db 1862 CATAGAGAGCTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1921
QY 1738 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1797
Db 1922 CAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1981
QY 1798 GATGTTTCAACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857
Db 1982 GATGTTTCAACACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2041
QY 1858 AAGCAGAGGTCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917
Db 2042 AAGCAGAGGTCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2101
QY 1918 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1940
Db 2102 GAGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2124

```

```

RESULT 3
CA324450 818 bp mRNA linear EST 26-NOV-2002
LOCUS UI-M-FY0-ccl-a-05-0-UI_r1 NIH BMAP_FY0 Mus musculus cDNA clone
DEFINITION IMAGE: 6820974 5', mRNA sequence.
ACCESSION CA324450.1 GI:24542548
VERSION CA324450.1
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 818)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
The following repetitive elements were found in this cDNA
sequence: 338-455, >(GGA)n#Simple_repeat
Seq primer: PYX-5.
FEATURES
SOURCE Location/Qualifiers
1..818
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6820974"
/class_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH BMAP FY0"
/notes="Organ: Brain; Vector: PYX-Anc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into PYX-Anc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCAGAGAG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
Program coordinator."
BASE COUNT 238 a 179 c 270 g 127 t 4 others
ORIGIN
Query Match 22.2%; Score 809.8; DB 14; Length 818;
Best Local Similarity 99.3%; Pred. No. 2e-171;
Matches 811; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 829 GAAGATTTATCTAGATGATGACCTAATGATGTCATGTCAGCCACTGCGC 888
Db 2 GAAGATTTATCTAGATGATGACCTAATGATGTCATGTCAGCCACTGCGC 61
QY 889 CCCGCTGCTCCGAATATCATCCCTCTCTTCCAGAAAGTTCCGTCAGATGCGG 948
Db 62 CCCGCTGCTCCGAATATCATCCCTCTCTTCCAGAAAGTTCCGTCAGATGCGG 121
QY 949 ATGTCCGTCATTAAGCTCTTCTGTGGATCAGAGGCGCTGAGAGCGGTCGTCAGAG 1008
Db 122 ATGTCCGTCATTAAGCTCTTCTGTGGATCAGAGGCGCTGAGAGCGGTCGTCAGAG 181
QY 1009 GATGAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1068
Db 182 GATGAGCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
QY 1069 TTGAGCAGCAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1128
Db 242 TTGAGCAGCAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 301
QY 1129 AAAGAGCAGAGAGCGGTTGGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAGAGAGAGAG 1188

```


Db 900 ACCCACTT 907

RESULT 5
AK082606
LOCUS
DEFINITION
AK082606 2589 bp mRNA linear HTC 05-DEC-2002
Mus musculus 0 day neonate cerebellar cDNA, RIKEN full-length
enriched library, clone:CC230068L04 product:intersectin (SH3 domain
protein 1A), full insert sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AK082606
AK082606.1 GI:26349812
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
99279253
10349636
Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20499374
11042159
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sun, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Yunoki, Y., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, U., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamada, I.,
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
Kuehl, P., Lewis, S., Matsumoto, Y., Nishida, I., Pesole, G.,
Quekkehueh, U., Schmitt, L. M., Staudt, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Flotcher, C., Fujita, M., Gariboldi, M., Guinacich, S., Hill, D.,
Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P.,
Rising, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L.,
Wyman-Borja, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohlschki, S.,
and Hayashizaki, Y.
Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11217851
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
6 (bases 1 to 2589)
Adachi, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Maruyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numata, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, K., Satoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M., and Hayashizaki, Y.
Direct Submission
Submitted (15-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp).
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1..2589
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:CC230068L04"
/db_xref="taxon:10090"
/clone="CC230068L04"
/clone="CC230068L04"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
/note="unlabeled protein product; intersectin (SH3 domain
protein 1A) (MGDI|GI:1338069, GB|NM_010587, evidence:
BLASTN 99%, match=2267)
putative"
/codon_start=2
/protein_id="BAC38546.1"
/translation="PHERGVRKSTSIDTGPTEPSALRKVAPPAKPAIAGEFIAM
VTYSSHQGLTFQGGVIVTKKDMGMGTGDKSGVPSNVYRLKDSGSGTAGC
TSGSGKRPETIAQVIAAAGPBLTARPOLIIRKNGGMEGLDARGKROIC
WFPANVYKLSPGSKITPTLPETIAQVAPACVIGVYDTAQNDDSLARSGQIIV
LNKEDPVMWGEVSGVGLPSPNVKLTMDPQVCSDLHLMTPPERKQGYI
HELIVTEENVNDQLVTEIFQKLTSELTBEVAMIFVNMELTMCNKLKAR
VRKMSGEKMPKVIIGILSAQLPHMYPYRFGSCQNGALIIQKTDPAEPFEVY
RLANDPRKMPKSSFLIKMQVRYTRPLIINKTENTPBPVPHSHKALAEAEEL
KQVNEVREKNSDRLEMTQAHVCGSLSEOLVPSVTNCTGPRKLHSGKLYKAS
NKEIVGLDFLLITITTPRSGSGTDYKPSKSNQYKTKPTPLNRYVLPD
PSGSEIFHSHIDRVYTLAESINERTAWQKKAASELITETKREKREAVYVRSQ
RATIGLWVNVGELIKPCRSKGNPNPCVEYMGQCHITKTIQTLNPKNSNCQ
PFIDLEOVNCTVFERDQLSADDFGRTEIRVADIKKQSGKSPVTKILLHVEPT
GEIVRDLQFDEP"

CDS
693 a 662 c 688 g 546 t

BASE COUNT
ORIGIN
Query Match 20.6%; Score 748.8; DB 11; Length 2589;
Best Local Similarity 99.7%; Pred. No. 1.8e-157;
Matches 750; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DB 2890 GGAGCCGTAAGAAATCCACAGCATCGATACGAGCCCTACTGAAAGTCTGCTAGCTTA 2949
14 GGCCCGTAAGAAATCCACAGCATCGATACGAGCCCTACTGAAAGTCTGCTAGCTTA 73

Db 481 ATGACAGACCCAGAGAAAGAACTGCTGTGACATTTGAAGATAAGAGCGGAGAACT 540
 QY 1070 TCGACGAGGCGCTGTGAGCTGAGAGACCGCCGCAAGCGTCTTTGAGCAGCAGCCCA 1129
 Db 541 TCGACGAGGCGCTGTGAGCTGAGAGACCGCCGCAAGCGTCTTTGAGCAGCAGCCCA 600
 QY 1130 AAGACAGAGAGCGTGTGCTGCTGAGCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAG 1189
 Db 601 AAGACAGAGAGCGTGTGCTGCTGAGCGCGCCGAGAGAGAGAGAGAGAGAGAGAGAG 660
 QY 1190 GCCAG 1249
 Db 661 GCCAG 720
 QY 1250 AGCTGAGAGCGG 1261
 Db 721 AGCTGAGAGCGG 732

RESULT 7
 CB248849 725 bp mRNA linear EST 14-FEB-2003
 LOCUS UI-M-EX0-by-m-24-0-UI.r1 NIH_BMAP_EX0 Mus musculus cDNA clone
 DEFINITION IMAGE: 5719103 5', mRNA sequence.
 CB248849
 VERSION CB248849.1 GI:28387663
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgs.nci.nih.gov/
 1 (bases 1 to 725)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgsabers@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILN at:
 http://image.llnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES

source
 Location/Qualifiers
 1..725
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 5719103"
 /tissue_type="whole brain"
 /dev_stage="embryo 15.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_idb="NIH_BMAP_EX0"
 /note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured mRNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with an
 oligo-dT primer containing a Not I site. Double stranded
 cDNA was size selected according to mRNA size fraction,
 ligated with EcoR I adaptor, digested with Not I, and then
 cloned directionally into pYX-Aac vector. The library tag
 sequence located between the Not I site and the polyA tail
 is GTGGCTGGA. This library was created for the
 University of Iowa Mouse Brain Molecular Anatomy Project
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous
 System', supported by National Institutes of Mental Health
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 250 a 151 c 221 g 102 t 1 others
 ORIGIN

Query Match 19.9%; Score 724; DB 14; Length 725;
 Best Local Similarity .99.9%; Pred. No. 4.1e-152;
 Matches 724; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1336 CTTGAATGGGAG 1395
 Db 1 CTTGAATGGGAG 60
 QY 1396 GGCACCGTGTCTTGAAG 1455
 Db 61 GGCACCGTGTCTTGAAG 120
 QY 1456 GACAAAAAGCATCAGCTAG 1515
 Db 121 GACAAAAAGCATCAGCTAG 180
 QY 1516 AGGCAAGAAATTGAG 1575
 Db 181 AGGCAAGAAATTGAG 240
 QY 1576 TTACAG 1635
 Db 241 TTACAG 300
 QY 1636 ATACTCAGTACCGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695
 Db 301 ATACTCAGTACCGCTTAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 QY 1696 ACCCTCAAAAGAGCTTGAAG 1755
 Db 361 ACCCTCAAAAGAGCTTGAAG 420
 QY 1756 GACGAGGTGAG 1815
 Db 421 GACGAGGTGAG 480
 QY 1816 CTGAAG 1875
 Db 481 CTGAAG 540
 QY 1876 GCAGGCGCAGTGAAG 1935
 Db 541 GCAGGCGCAGTGAAG 600
 QY 1936 GAAAGCGCTCAG 1995
 Db 601 GAAAGCGCTCAG 660
 QY 1996 GAG 2055
 Db 661 GAG 720
 QY 2056 GTCAG 2060
 Db 721 GTCAG 725

RESULT 8
 BM950476 732 bp mRNA linear EST 14-MAR-2002
 LOCUS UI-M-EH0P-buu-o-05-0-UI.r1 NIH_BMAP_EH0P Mus musculus cDNA clone
 DEFINITION IMAGE:5687260 5', mRNA sequence.
 BM950476
 VERSION BM950476.1 GI:19434066
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgs.nci.nih.gov/
 1 (bases 1 to 732)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
 source
 Location/Qualifiers
 1.738
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5687260"
 /tissue_type="whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP EH0P"
 /note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System, supported by National Institute of Mental Health (NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT 208 a 172 c 201 g 148 t 3 others

Query Match 19.8%; Score 722.8; DB 12; Length 732;
 Best Local Similarity 99.3%; Pred. No. 7.6e-152;
 Matches 724; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

2817 AGACATGTGTGTTGAGAGGTTCAAGGTCAGAGGTTGTTCCCAAGTCTTACGT 2876
 4 ACATNNGTGTGTGTTGAGAGGTTCAAGGTCAGAGGTTGTTCCCAAGTCTTACGT 63
 2877 GAAACTCATTTACAGGGCCGTAAGAAATCCACAGCATGATCTGCGCTTACTGAAG 2936
 64 GAAACTCATTTACAGGGCCGTAAGAAATCCACAGCATGATCTGCGCTTACTGAAG 123
 2937 TCCTGCTAGCTTAAAGAGATGCTTCCCGCGCCCAACCGCATTTCCCGAGAGA 2996
 124 TCCTGCTAGCTTAAAGAGATGCTTCCCGCGCCCAACCGCATTTCCCGAGAGA 183
 2997 GTTATATGCATGATACATACAGAGATGCTGAGCAAGAGATTTAACTTTACAGCAAG 3056
 184 GTTATATGCATGATACATACAGAGATGCTGAGCAAGAGATTTAACTTTACAGCAAG 243
 3057 GAGATGATTTGTGTTACCAAGAAAGATGTGACTGTGACGCGAAACGCTGGCGCA 3116
 244 GAGATGATTTGTGTTACCAAGAAAGATGTGACTGTGACGCGAAACGCTGGCGCA 303
 3117 GTCCGAGGTTCTCCCTTCTAACTATGTAGGCTTAAAGATTCAGAGGCTCTGGAATGC 3176
 304 GTCCGAGGTTCTCCCTTCTAACTATGTAGGCTTAAAGATTCAGAGGCTCTGGAATGC 363
 3177 TGGGAAAAACAGGAGTTTAAAGAAAAAAGTGAATGCGGATTTGCTTCTTACGC 3236

Db 364 TGGGAAAAACAGGAGTTTAAAGAAAAAAGTGAATGCGGATTTGCTTCTTACGC 423
 3237 TGCTATCTGTGTTCCCAAGCAATCACTCCCTGCTCTGCGACGCTGATTCGCGAAAA 3296
 424 TGCTATCTGTGTTCCCAAGCAATCACTCCCTGCTCTGCGACGCTGATTCGCGAAAA 483
 3297 GAACCCAGGTTGATGTGGAGAGAGAACTGCAAGTCCAGGGAAGAACCGCAGATAG 3356
 484 GAACCCAGGTTGATGTGGAGAGAGAACTGCAAGTCCAGGGAAGAACCGCAGATAG 543
 3357 GTGTTTCCAGCAAAATTTATGCAAACTTCAAGCCCGGAGCAAGCAAAATCAACCCCA 3416
 544 GTGTTTCCAGCAAAATTTATGCAAACTTCAAGCCCGGAGCAAGCAAAATCAACCCCA 603
 3417 TGAGCTAACCAAGACCGCAGTGCAGCAGCAGTGTGCGAGTGATCGGATGTACATTA 3476
 604 TGAGCTAACCAAGACCGCAGTGCAGCAGCAGTGTGCGAGTGATCGGATGTACATTA 663
 3477 CACCGCCAGAACGATGAGAACTGACCTTCAGCAAGGCGCAGATCATCAAGCTCTCA 3536
 664 CACCGCCAGAACGATGAGAACTGACCTTCAGCAAGGCGCAGATCATCAAGCTCTCA 723
 3537 CAAGAGAGA 3545
 Db 724 CAAGAGAGA 732

RESULT 9
 BM944544 738 bp mRNA linear EST 14-MAR-2002
 LOCUS U1-M-EH0P-pyr-j-08-0-UI r1 NIH_BMAP_EH0P Mus musculus cDNA clone
 DEFINITION IMAGE:5695975 5', mRNA sequence.
 ACCESSION BM944544 GI:19428129
 VERSION BM944544
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/HLN at: <http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
 source
 Location/Qualifiers
 1.738
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:5695975"
 /tissue_type="whole brain"
 /dev_stage="embryo 18.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_1ib="NIH BMAP EH0P"
 /note="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded

University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System," supported by National Institute of Mental Health (NIMH), Hemlin Chiu, Ph.D., program coordinator."

BASE COUNT	198 a	213 c	184 g	143 t
ORIGIN				

Query Match	19.8%	Score 721.2;	DB 12;	Length 738;
Best Local Similarity	99.5%;	Pred. No. 1.8e-151;		
Matches 734;	Conservative 0;	Mismatches 3;	Indels 1;	Gaps 1;

QY	2310	CCAGCTGTGAGAGCCAGGATGGCTTGGAGGAGCTGAAAGGGAAGACGGGATGGTTCC	236
Db	1	CCAGCTGTGAGAGCCAGGATGGCTTGGAGGAGCTGAAAGGGAAGACGGGATGGTTCC	60
QY	2370	TGCAAACTATGCAGAAAAGATTCCAGAAAATGAGGTTCCCATCTCCAGCCAAACCGTAGAC	242
Db	61	TGCAAACTATGCAGAAAAGATTCCAGAAAATGAGGTTCCCATCTCCAGCCAAACCGTAGAC	120
QY	2430	CGATTTGACATCTGCCCCCTGCCCCAAACTGGCTTTGCGTGAAGACCCCTGCTCTTTGCC	248
Db	121	CGATTTGACATCTGCCCCCTGCCCCAAACTGGCTTTGCGTGAAGACCCCTGCTCTTTGCC	180
QY	2490	AGTAGCCCTTCTGAGGCCCTCCACAAACCCCAACAACTGGGAGACTTCAGTTCCACGTC	254
Db	181	AGTAGCCCTTCTGAGGCCCTCCACAAACCCCAACAACTGGGAGACTTCAGTTCCACGTC	240
QY	2550	GCCGAGCAGCTCAAAACGAGAACCGAACGCAACTCGGATACGTGGCGGCTCAAGCC	260
Db	241	GCCGAGCAGCTCAAAACGAGAACCGAACGCAACTCGGATACGTGGCGGCTCAAGCC	300
QY	2610	TTCTCTGACCGTACCTAAGTGTGCTGSCAAGTTACGCGAGATCAGCCTTTACCCACGAC	266
Db	301	TTCTCTGACCGTACCTAAGTGTGCTGSCAAGTTACGCGAGATCAGCCTTTACCCACGAC	360
QY	2670	AGGCACGTGGCTCTCCGCCATGCTCCCGTCTGGGCCAGGGGTGAAGAGGGCTTCA	272
Db	361	AGGCACGTGGCTCTCCGCCATGCTCCCGTCTGGGCCAGGGGTGAAGAGGGCTTCA	420
QY	2730	AGCGCAAGCCCTGTATCCCTGGAGAGCCAAAAAACACACACTTAAATTTTAAACAAAG	278
Db	421	AGCGCAAGCCCTGTATCCCTGGAGAGCCAAAAAACACACACTTAAATTTTAAACAAAG	480
QY	2790	TGACGTATCACCCGTTCTTGAAACAGCAAGACATGTGTGTGTTTGGAGAACTTCAAGTCA	284
Db	481	TGACGTATCACCCGTTCTTGAAACAGCAAGACATGTGTGTGTTTGGAGAACTTCAAGTCA	540
QY	2850	GAAAGGTTGGTTCCTCCAGTCTTACGTAAACATCTTTCAGGGCCCGTAAGAAATCCAC	290
Db	541	GAAAGGTTGGTTCCTCCAGTCTTACGTAAACATCTTTCAGGGCCCGTAAGAAATCCAC	600
QY	2910	AAGCATCATACTGGCCCTTACTGAAGAGTCTGCTAGTCTTAAAGAGATGGCTTCCCGAC	296
Db	601	AAGCATCATACTGGCCCTTACTGAAGAGTCTGCTAGTCTTAAAGAGATGGCTTCCCGAC	660
QY	2970	CGCCAGCCAGCCATTCCCGGAGAAAGATTATTCATGTACACATACGAGATTTCTGA	302
Db	661	CGCCAGCCAGCCATTCCCGGAGAAAG-TTATTCATGTATCACATACGAGATTTCTGA	719
QY	3030	GCAAGGAGATTTAACCTT 3047	
Db	720	GCAAGGAGATTTAACCTT 737	

RESULT 10	LOCUS	DEFINITION	729 bp	mRNA	linear	EST 09-OCT-2007
BU704308	BU704308	UT-M-FOO-b2s-j-21-0-UT.r1 NIH_BMAP_F00 Mus musculus cDNA clone				

IMAGE: 6406772 5', mRNA sequence.

ACCESSION	BU704308
RECTION	BU704308 1 CT-33633308

KEYWORDS EST.

SOURCE	Mus musculus (mouse)
ORGANISM	Mus musculus

Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Rodentia

REFERENCE 1 (bases 1 to 729)

TITLE National Institutes of Health

COMMENT Contact: Robert Strausberg

Email: cgyabbs-remail.hin.s
Tissue Procurement: Dr. Ji

CDNA Library preparation:
CDNA libraries were prepared by

DNA Sequencing by: Dr. M.

found through the I.M.A.G.

<http://image.11n1.gov>
This clone was contributed

(BMAP)
FOR DISMANTLING

FEATURES

Location/Qualifie

1. 1/23
/organism="Mus mu

```
/mol_type="mRNA"  
/strain="C57BL/6"
```

```
/db_xref="taxon:107390"IMCE: 64
```

/tissue_type="whc

```
/lab_host="DH10B"
```

```

/clone_11b="N1H_E
/note="Organ: Brä

```

Site 2: Not I; Th

1996. Denatured F

primer containing

size selected acc
with EcoR I adapt

directionally into
sequence located

is TGAGAGACC. The

Developing Mouse

Program Coordinator
Institute of Mental Health

BASE COUNT	176 a	206 c	174
------------	-------	-------	-----

Query match	45.38%	Score
Best Local Similarity	99.28%	Predicted

Matches 719; Conservative 0; N

382 GCTGTTGCTCCTGTGCCAATGGG

Db 1 GCTGTTGCTCCTGTGCCAATGGG

QY 442 TCTTCTGTCCCTCCAGCAGCAGTC

Db	61	TCTTGTGCTCCGAGCAGAGTGCTCCCTGGGCTTAACGGGGCTTCCTCCGTCATACAG	120
QY	502	CCTCGCTGGGTTGGCGATCTCGACGCACATGGCCAAAGTTCTTCTTCAGCAAA	561
Db	121	CCTCGCTGGGTTGGCGCATCTCT--GCACATTGCCAAAGTCTTCTTCAGCAAA	177
QY	562	TCTGGTCCAGGCTCACTTAACACTAAGTATGAGAAAGGCACATCATTCGATGCGCC	621
Db	178	TCTGGTCCAGGCTCACTTAACACTAAGTATGAGAAAGGCACATCATTCGATGCGCC	237

```

QY 622 AGGCCCCCTCAGCAGAGAGATGGGCTGTGCTCACTCATCAAGCTGAATACAGCAG 681
DB 238 AGGCCCCCTCAGCAGAGAGATGGGCTGTGCTCACTCATCAAGCTGAATACAGCAG 297
QY 682 TTATTCACAGCAGCAGCAAAACTATGATGACACTTAAACAGGTCCCAAGCAAGT 741
DB 298 TTATTCACAGCAGCAGCAAAACTATGATGACACTTAAACAGGTCCCAAGCAAGT 357
QY 742 ATTCTCATGATCAATCAAGTTTAACTCCAGGCTCAGTGTCTTCAATATGAAATCTTTCTGAC 801
DB 358 ATTCTCATGATCAATCAAGTTTAACTCCAGGCTCAGTGTCTTCAATATGAAATCTTTCTGAC 417
QY 802 ATTGATCAAGATGAAATCACTCATGAGAGAAATTTATCTAGTATGACCTAATGAT 861
DB 418 ATTGATCAAGATGAAATCACTCATGAGAGAAATTTATCTAGTATGACCTAATGAT 477
QY 862 GTTGCCATGTCTGTGTCAGCAGCAGTCCGCGCTCTGCTCAGAAATACATCCCTCTCC 921
DB 478 GTTGCCATGTCTGTGTCAGCAGCAGTCCGCGCTCTGCTCAGAAATACATCCCTCTCC 537
QY 922 TTGCAAGAGTTGCTGCTCCGCACTGGATGGATCCGCTCATAGCTCTTCTGTGATCAG 981
DB 538 TTGCAAGAGTTGCTGCTCCGCACTGGATGGATCCGCTCATAGCTCTTCTGTGATCAG 597
QY 982 AGGCTGCTGAGAGAGCCGCTGCTGATGAGATGAGCAGCAGCAGAAAGAACTGCTGTG 1041
DB 598 AGGCTGCTGAGAGAGCCGCTGCTGATGAGATGAGCAGCAGCAGAAAGAACTGCTGTG 657
QY 1042 ACATTTGAAGATTAAGAGCGGAGAACTTTCAGAGCAGCAGTGTGAGCTGAGAAAGCG 1101
DB 658 ACATTTGAAGATTAAGAGCGGAGAACTTTCAGAGCAGCAGTGTGAGCTGAGAAAGCG 717
QY 1102 CGCCA 1106
DB 718 CGCAA 722

RESULT 11
LOCUS CBS21237 706 bp mRNA linear EST 28-MAR-2003
DEFINITION U1-M-GH0-cem-g-10-0-U1.r1 NIH BMAP_GH0 Mus musculus cDNA clone
IMAGE: 6841475 5', mRNA sequence.
ACCESSION CBS21237
VERSION CBS21237.1 GI:29354592
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
MUSCULUS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 706)
NIH-MGC http://mgc.ncl.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: rs9abds-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.iowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP).
The following repetitive elements were found in this cDNA
sequence: 220-337, >(GGA)n#simple_repeat
Seq primer: pyx-5.
FEATURES
Source
1..706
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6841475"

```

```

/tissue_type="Whole brain"
/dev_stage="1, 5, and 15 days newborn"
/lab_host="DH10B (T1 phage resistant)"
/clone_idb="NIH BMAP_GH0"
/notes="Organ: Brain; Vector: pyx-Asc; Site 1: Ecor I;
Site 2: Not I; The library was constructed according
to Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mDNA size fraction, ligated
with Ecor I adaptor, digested with NotI and then cloned
directionally into pyx-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is CCAACTGAGT. This library was created for the University
of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the
Developing Mouse Nervous System", supported by National
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
program coordinator."
BASE COUNT 219 a 143 c 247 g 96 t 1 others
ORIGIN
Query Match 19.3%; Score 701.8; DB 14; Length 706;
Best Local Similarity 99.6%; Pred. No. 4e-147;
Matches 703; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 946 GGGATGTCGCTCATTAAGCTTCTTCTGTGATCAGAGGCTGCTAGAGCCGTGTCA 1005
DB 1 GGGATGTCGCTCATTAAGCTTCTTCTGTGATCAGAGGCTGCTAGAGCCGTGTCA 60
QY 1006 GAGGATGACAGAGCAGAGCAAGAACTGCTGTGATCAAGAGCTGAGAGCCGTGTCA 1065
DB 61 GAGGATGACAGAGCAGAGCAAGAACTGCTGTGATCAAGAGCTGAGAGCCGTGTCA 120
QY 1066 AACTTGACGAGGAGCAGTGTGAGTGAAGAGCCGCGCAAGCGCTTGTGAGCAGAG 1125
DB 121 AACTTGACGAGGAGCAGTGTGAGTGAAGAGCCGCGCAAGCGCTTGTGAGCAGAG 180
QY 1126 CGCAAGAGCAGAGCGGTTGCTCAGCTGAGCGCGCGAGAGAGAGAGAAAGCGG 1185
DB 181 CGCAAGAGCAGAGCGGTTGCTCAGCTGAGCGCGCGAGAGAGAGAGAAAGCGG 240
QY 1186 GAGCGCAGAGAGCAGAGCGGAGCGGAGCTGAGTGAAGAGCAGCTGAGAGAGCAG 1245
DB 241 GAGCGCAGAGAGCAGAGCGGAGCGGAGCTGAGTGAAGAGCAGCTGAGAGAGCAG 300
QY 1246 CGGAGCTGAGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
DB 301 CGGAGCTGAGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 1306 GCAAAACGGAACTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1365
DB 361 GCAAAACGGAACTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 1366 CTCCTGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
DB 421 CTCCTGAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 1426 ACTCTGAGATTTGAGTTAGTGAAGCTGGAATGACAAAACATCAGCTAGAGAGAACTT 1485
DB 481 ACTCTGAGATTTGAGTTAGTGAAGCTGGAATGACAAAACATCAGCTAGAGAGAACTT 540
QY 1486 CAGGATATCAGGTGTGACTGCGCAACCCAGAGGCAAGAAATTTAGAGCACGAAGTCT 1545
DB 541 CAGGATATCAGGTGTGACTGCGCAACCCAGAGGCAAGAAATTTAGAGCACGAAGTCT 600
QY 1546 AGAGACTAAGATTGCTGAATACCCAGCTTACAGCAGAGTTGAGAGAGATTCAGCA 1605
DB 601 AGAGACTAAGATTGCTGAATACCCAGCTTACAGCAGAGTTGAGAGAGATTCAGCA 660
QY 1606 ATGCTTGAAGACTTATTCAGAGAAACAGATACCTAGTACAGT 1651
DB 661 ATGCTTGAAGACTTATTCAGAGAAACAGATACCTAGTACAGT 706

```

RESULT 12
 CBS27154 703 bp mRNA linear EST 28-MAR-2003
 LOCUS UI-M-FY0-cdc-j-23-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6849689 5', mRNA sequence.
 CBS27154
 ACCESSION CBS27154.1 GI:29360627
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 703)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbe-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at <http://genome.uiowa.edu/distribution/mousefl.html>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: PYX-5.
 Location/Qualifiers
 1..703
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE: 6849689"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_lib="NIH_BMAP_FY0"
 /note="Organ: Brain; Vector: PYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is ACCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chan, Ph.D., program coordinator."

BASE COUNT 228 a 154 c 225 g 94 t 2 others
 ORIGIN

Query Match 19.2%; Score 699.4; DB 14; Length 703;
 Best Local Similarity 99.6%; Pred. No. 1.4e-146;
 Matches 700; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1699 CTCAAAAGACCTTGGAGCAAGAGAGCTGCGCCGCGACAGCTCCGGAGAGAGCTGGAC 1758
 DB 1 CTCAAAAGAGCCTTGGAGCAAGAGAGCTGCGCCGCGACAGCTCCGGAGAGAGCTGGAC 60
 QY 1759 GAGGTGAG 1818
 DB 61 GAGGTGAG 120
 QY 1819 AAGGAAGTGAAG 1878
 DB 121 AAGGAAGTGAAG 180

QY 1879 GCGGACTGAG 1938
 DB 181 GCGGACTGAG 240
 QY 1939 GAGGCTGAG 1998
 DB 241 GAGGCTGAG 300
 QY 1999 GAGGAGCCAG 2058
 DB 301 GAGGAGCCAG 360
 QY 2059 AGGAAG 2118
 DB 361 AGGAAG 420
 QY 2119 TTCATCCGATCAG 2178
 DB 421 TTCATCCGATCAG 480
 QY 2179 AAAGGCGCGCTTACCATTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2238
 DB 481 AAAGGCGCGCTTACCATTTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
 QY 2239 TACCCCTTGAATCAG 2298
 DB 541 TACCCCTTGAATCAG 600
 QY 2299 GTGATGAAAGCCAG 2358
 DB 601 GTGATGAAAGCCAG 660
 QY 2359 GGATGCTTCCCTGCAAACTATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2401
 DB 661 GGATGCTTCCCTGCAAACTATGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703

RESULT 13
 CBS29050 713 bp mRNA linear EST 27-NOV-2002
 LOCUS UI-M-FY0-cdc-j-23-0-UI.r1 NIH BMAP_FY0 Mus musculus cDNA clone
 DEFINITION IMAGE: 6827736 5', mRNA sequence.
 CBS29050
 ACCESSION CBS29050.1 GI:24547148
 VERSION EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 713)
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbe-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 The following repetitive elements were found in this cDNA sequence: 341-458, >(GGA)n#Simple_repeat
 Seq primer: PYX-5.
 Location/Qualifiers
 1..713
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"

```

/db_xref="taxon:10090"
/clone="IMAGE: 6827736"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dp"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."
BASE COUNT      199 a      157 c      245 g      106 t      6 others
ORIGIN

```

```

Query Match      19.0%; Score 691.2; DB 14; Length 713;
Best Local Similarity 98.6%; Pred. No. 9,9e-145;
Matches 704; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 825 TGCAGAAAGATTATCTTACTTACTTATGCACTTAATTGATGTCATGTCGACGCACT 884
    1 TGCAGAAAGATTATCTTACTTACTTATGCACTTAATTGATGTCATGTCGACGCACT 60
DB 1 GCGCCCGCTCTGCTCCAGATATACCTCTCTCTCTGAGAAAGTTGCTCCGCGAG 944
    61 GCGCCCGCTCTGCTCCAGATATACCTCTCTCTCTGAGAAAGTTGCTCCGCGAG 120
QY 945 TGGGATGTCGCTGATTAAGCTCTTCTTCTGATGTCAGAGGCTCCGTCGTCGTC 1004
    121 TGGGATGTCGCTGATTAAGCTCTTCTTCTGATGTCAGAGGCTCCGTCGTCGTC 180
DB 1005 AGAGATGACAGACAGCCAGAGAAAGAACTGCTGTGACATTTGAAGTAAGAGCGGAG 1064
    181 AGAGATGACAGACAGCCAGAGAAAGAACTGCTGTGACATTTGAAGTAAGAGCGGAG 240
QY 1065 GAACCTGAGCGGAGGAGCTGTGAGTGTGAGAGAGCGCCGCAAGCCCTCTTGAGAGCA 1124
    241 GAACCTGAGCGGAGGAGCTGTGAGTGTGAGAGAGCGCCGCAAGCCCTCTTGAGAGCA 300
DB 1125 GCGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGCG 1184
    301 GCGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGAGCAGAGAGAGAGAGCG 360
QY 1185 GAGCGCCAGAGCAGAGCGGCAAGCGGAGCTGTGAGAGCAGAGCAGCTGTGAGAGCA 1244
    361 GAGCGCCAGAGCAGAGCGGCAAGCGGAGCTGTGAGAGCAGAGCAGCTGTGAGAGCA 420
DB 1245 GCGGAGCTGTGAGCGGAGCAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAGAGC 1304
    421 GCGGAGCTGTGAGCGGAGCAGAGAGAGAGAGAGAGATGAGAGCGCGCAGAGC 480
QY 1305 CGCAAAACGGGAACCTGAAAGCAGCAACTTGAATGGGAACGGAACCGAGACAGGA 1364
    481 CGCAAAACGGGAACCTGAAAGCAGCAACTTGAATGGGAACGGAACCGAGACAGGA 540
DB 1365 ACTCTGGAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1424
    541 ACTCTGGAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 1425 GACTCTGAGGTTGAGTGAAGCTCTGATGCAAAAGAGCAGCAGTGAAGAGAGAGAG 1484
    601 GACTCTGAGGTTGAGTGAAGCTCTGATGCAAAAGAGCAGCAGTGAAGAGAGAGAG 659
DB 1485 TCAGATATCAGGTTGCTGAGCAACCCAGAGCAGAAAGTAATGAGAGCAGGA 1538

```

Db 660 TCAGATATCAGGTTGCTGAGCAACCCAGAGAGAGAAATTGAGAGCAGGA 713

```

RESULT 14
CB244555
LOCUS
DEFINITION
CB244555
ACCESSION
CB244555
VERSION
CB244555.1
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NIH-MGC http://mgi.mc.man.ac.uk/
NIH-MGC http://mgi.mc.man.ac.uk/
Unpublished
Contact: Robert Strauberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
sequence: 413-530, >(GGA)nSimple_repeat
Seq primer: pYX-5.
Location/Qualifiers
1..688
/organism="Mus musculus"
/mol_type="cDNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6833270"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dp"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FY0"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."
BASE COUNT      188 a      156 c      235 g      109 t
ORIGIN

```

```

Query Match      18.8%; Score 684.8; DB 14; Length 688;
Best Local Similarity 99.7%; Pred. No. 2.7e-143;
Matches 686; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 753 ATCAAGTTACCCAGGCTGCTGCTTCAATATGAAATCTTTGACATTGATCAAGA 812
    1 ATCAAGTTACCCAGGCTGCTGCTTCAATATGAAATCTTTGACATTGATCAAGA 60
DB 813 TGAAGAACTCACTGAGAGAAATTTACTTACTTATGACCTAATTGATGTCATGTC 872

```

```

Db      61 TGGAAAGTCACTGAGAGAAATTTATCTAGCTATGACCTATATGATGTTGCATGTC 120
QY      873 TGGTACGACACTGCGCGCCGCTCTGCTTCACAAATACATCCCTCTTCTTCAGAAAGT 932
Db      121 TGGTACGACACTGCGCGCCGCTCTGCTTCACAAATACATCCCTCTTCTTCAGAAAGT 180
QY      933 TGGCTCCGAGTGGGATGTCCTGATCAAGTCTTCTTCTGATCAAGAGGCTGCTGTA 992
Db      181 TGGCTCCGAGTGGGATGTCCTGATCAAGTCTTCTTCTGATCAAGAGGCTGCTGTA 240
QY      993 GAGGCGCTGTCAGAGATGAGCAGACGCGAGAGAAAGTCTGTCATTTGAAGA 1052
Db      241 GAGGCGCTGTCAGAGATGAGCAGACGCGAGAGAAAGTCTGTCATTTGAAGA 300
QY      1053 TAAAGACCGGAGAACTTTCAGACGAGGCTGAGGCTGAGAAAGCGCCGAGCGCT 1112
Db      301 TAAAGACCGGAGAACTTTCAGACGAGGCTGAGGCTGAGAAAGCGCCGAGCGCT 360
QY      1113 CTTCGAGCAGACGCGAGAGAGAGAGGCTGTCAGTTCAGAGCGCGCGAGCAGGA 1172
Db      361 CTTCGAGCAGACGCGAGAGAGAGAGGCTGTCAGTTCAGAGCGCGCGAGCAGGA 420
QY      1173 GAGGAGAGAGCGGAGCGCGCAGAGAGAGAGCGCAAGCGCGAGCTGAGCTGAGAGCA 1232
Db      421 GAGGAGAGAGCGGAGCGCGCAGAGAGAGAGCGCAAGCGCGAGCTGAGCTGAGAGCA 480
QY      1233 GCTGAGAGAGCGGAGCGCGAGCTGAGCGGAGAGAGAGAGAGAGAGAGATGCA 1292
Db      481 GCTGAGAGAGCGGAGCGCGAGCTGAGCGGAGAGAGAGAGAGAGAGAGATGCA 540
QY      1293 GAGGCGCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1352
Db      541 GAGGCGCGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY      1353 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1412
Db      601 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY      1413 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db      661 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 688

RESULT 15
LOCUS   BQ719508 979 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT 8219793 Lupski symptomatic_trunk Homo sapiens cDNA clone
IMAGE:6188111 5', mRNA sequence.
ACCESSION BQ719508
VERSION   .BQ719508.1 GI:21858405
KEYWORDS EST.
SOURCE    Homo sapiens (human)
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 979)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLES    Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: csgrabs@mail.nih.gov
           Tissue Procurement: Dr. James R. Lupski
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LHAM13583 row: c column: 24
           High quality sequence stop: 643.
           Location/Qualifiers
             1..979
           /organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6188111"
/sex="male"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/clone_1b="Lupski symptomatic trunk"
/notes="Vector: pCMV-Sport6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCG-3' and
5'-GACTAGTCTAGATCGAGCGGCGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

```

```

BASE COUNT 294 a 247 c 218 g 217 t 3 others
ORIGIN

```

```

Query Match 18.1%; Score 658.2; DB 13; Length 979;
Best Local Similarity 84.2%; Pred. No. 3.1e-137;
Matches 776; Conservative 0; Mismatches 141; Indels 5; Gaps 3;

```

```

QY      291 ACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCAGATGACACGATTTGAT 350
Db      1 ACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCAGATGACACGATTTGAT 60
QY      351 AGAGAGAGATGTCATGATGACACCACTACAGCTGTTGCTCTGTGCAATGGCTCAT 410
Db      61 GAGAGATATGCGCAGATGACACCGCTTACAGCTGTTGCTCTGTGCAATGGATCAT 120
QY      411 TCCAGTTGTTGATGTCCTCACCCTTAATGTCCTTCTGCTCTCCACAGAGTCTCC 470
Db      121 TCCAGTTGTTGATGTCCTCACCCTTAATGTCCTTCTGCTCTCCACAGAGTCTCC 180
QY      471 CCTGAGTAAAGGAGGCTCTCCCGCATACAGCCCTGCTGCTGCTGCTGCTGAGC 530
Db      181 CCTGAGTAAAGGAGGCTCTCCCGCATACAGCCCTGCTGCTGCTGCTGCTGAGC 240
QY      531 CACATGGCCAAAGATTCTTCTTACAGAGATGTCACAGGTCACAGATTAACACTAA 590
Db      241 CACATGGCCAAAGATTCTTCTTACAGAGATGTCACAGGTCACAGATTAACACTAA 300
QY      591 GTTACAGAGGCAATATTCATGTCGCGCAGCGCCCTTCAGACAGAAATGGCTGT 650
Db      301 ATTACAAAGGCAAGTATTCATGTCGCGCAGCGCCCTTCAGACAGAAATGGCTGT 360
QY      651 GCTCAGTCATCAAGGCTGAAATACAGGAGTTATTAACAGCAGCAAAACTATGAG 710
Db      361 TCTCAGTCATCAAGGCTGAAATACAGGAGTTATTAACAGCAGCAAAACTATGAG 420
QY      711 TGGACACTTAACAGAGTCCCGCAGAGAACTATTCATGCAATCAAGTTTACCCAGGC 770
Db      421 TGGACACTTAACAGAGTCCCGCAGAGAACTATTCATGCAATCAAGTTTACCCAGGC 480
QY      771 TCAGTGGCTCAATATGGAATCTTTTCTGACATTCATGATGAAATCACTGACGA 830
Db      481 TCAGTGGCTCAATATGGAATCTTTTCTGACATTCATGATGAAATCACTGACGA 540
QY      831 AGAATTTATCTAGCTATGACCTAATTTGATGTCGATGTCGTCAGACCTGCGCC 890
Db      541 GGAATTTATCTAGCTATGACCTAATTTGATGTCGATGTCGTCAGACCTGCGCC 600
QY      891 CGTCTGCTCCAGAAATCATCTCTCTCTTCAGAGAGTTGCTCCGCGAGTGGAT 950
Db      601 TGTCTGCTCCAGAAATCATCTCTCTCTTCAGAGAGTTGCTCCGCGAGTGGAT 660
QY      951 GTCTGCTATTAAGCTCTTCTTCTGATGATCAGAGGCTGTCAGAGAGCCGTGTCAGAGA 1010
Db      661 ATCTGCTATTAAGCTCAATCTGTATGATCAGAGGCTACAGAGAGAAAGATTGTAAGA 720

```

```

QY 1011 TGAGCAGCAGC---CAGAGAGAAAATGCTGTGACATTGAG-ATAAGAGCGGAGA 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TGAACAACAATTAAGAAAATTAACCTGTAACTTTGAAGATTAAGAGCGGAGA 780
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1067 ACTTCGAGCGAGGCACTGTGAGCTGAGAAAGCGCCCAAGCGCTCTTGAGCAGCAGC 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 ACTTTGAACGTGGCAACCTGGAAGTGAAGAAAGAG-CAAGCTCTCTGAGAAAGCGCAGC 839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1127 GCAAAAGAGCAGAGCGGTTGGCTCAGCTGAGAGCGCGCGGAGCAGAGAGAGAAAGCGCGG 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 840 GCAAGGAAGCAGAGCGCTGGCCANNCTGAACGAGCGGAGCAGAGAAAGAGAAAGT 899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1187 AGCGCCAGAGCAGAGAGGCCAA 1208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 900 GACCCCAAGAACAGAGCGCAAA 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: December 5, 2003, 00:39:55
 Job time : 4658.38 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 4, 2003, 15:25:00 ; Search time 158.185 Seconds
(without alignments)
10162.293 Million cell updates/sec

Title: US-09-674-237A-2
Perfect score: 3642
Sequence: 1 atggtcagcttcacccaccc.....tgaccacccagcaatga 3642

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued Patents NA:
2: /cgn2_6/prodata/2/ina/5A COMB. seq:
3: /cgn2_6/prodata/2/ina/5A COMB. seq:
4: /cgn2_6/prodata/2/ina/5A COMB. seq:
5: /cgn2_6/prodata/2/ina/5A COMB. seq:
6: /cgn2_6/prodata/2/ina/5A COMB. seq:
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	477.4	13.1	2017	US-09-404-879A-72	Sequence 72, App1
2	477.4	13.1	2017	US-09-338-933-72	Sequence 72, App1
3	477.4	13.1	2017	US-09-215-681-72	Sequence 72, App1
4	332.4	9.1	2873	US-08-630-915A-193	Sequence 193, App1
5	258	7.1	747	US-08-630-915A-193	Sequence 193, App1
6	144	4.0	480	US-09-404-879A-60	Sequence 60, App1
7	144	4.0	480	US-09-338-933-60	Sequence 60, App1
8	144	4.0	480	US-09-215-681-60	Sequence 60, App1
9	144	4.0	531	US-09-404-879A-5	Sequence 5, App1
10	144	4.0	531	US-09-338-933-5	Sequence 5, App1
11	144	4.0	531	US-09-215-681-5	Sequence 5, App1
12	144	4.0	3489	US-08-728-323A-1	Sequence 1, App1
13	144	4.0	3489	US-09-298-568-1	Sequence 1, App1
14	144	4.0	3489	US-09-410-399-1	Sequence 1, App1
15	144	4.0	32207	US-08-770-379-20	Sequence 20, App1
16	144	4.0	32207	US-08-757-669A-20	Sequence 20, App1
17	144	4.0	32207	US-09-230-371A-20	Sequence 20, App1
18	121.6	3.3	531	US-09-404-879A-4	Sequence 4, App1
19	121.6	3.3	531	US-09-338-933-4	Sequence 4, App1
20	121.6	3.3	531	US-09-215-681-4	Sequence 4, App1
21	119.6	3.3	9551	US-08-056-200-93	Sequence 93, App1
22	119.6	3.3	9551	US-08-800-644-93	Sequence 93, App1
23	119.2	3.3	1926	US-09-249-585A-2	Sequence 2, App1
24	119.2	3.3	1926	US-09-410-399-3	Sequence 3, App1
25	119.2	3.3	2580	US-09-050-863-2	Sequence 2, App1
26	119.2	3.3	2580	US-09-359-081-2	Sequence 2, App1
27	119.2	3.3	5452	US-09-130-114-1	Sequence 1, App1

C	28	119.2	3.3	8705	4	US-09-647-344A-14	Sequence 14, App1
	29	119.2	3.3	9600	3	US-08-910-647-1	Sequence 1, App1
	30	119.2	3.3	9600	4	US-09-620-925-1	Sequence 1, App1
	31	119.2	3.3	10596	1	US-07-884-811-15	Sequence 15, App1
	32	119.2	3.3	10596	1	US-07-885-971-15	Sequence 15, App1
	33	119.2	3.3	10596	1	US-08-087-783A-15	Sequence 15, App1
	34	119.2	3.3	10596	1	US-08-194-088B-15	Sequence 15, App1
	35	119.2	3.3	10596	1	US-08-194-087-15	Sequence 15, App1
	36	119.2	3.3	10596	5	PCR-US93-04648-15	Sequence 15, App1
	37	107	2.9	7218	1	US-08-232-463-14	Sequence 14, App1
	38	95.8	2.6	16442	3	US-08-781-891-208	Sequence 208, App1
	39	95.8	2.6	16442	4	US-09-618-166-208	Sequence 208, App1
	40	87.4	2.4	397	3	US-09-253-691-3	Sequence 3, App1
	41	87.4	2.4	51259	3	US-08-781-891-209	Sequence 209, App1
	42	87.4	2.4	51259	4	US-09-618-166-209	Sequence 209, App1
	43	86.2	2.4	543	6	5273901-6	Sequence 6, App1
	44	83.6	2.3	1037	4	US-09-181-585-3	Sequence 3, App1
	45	83.6	2.3	1159	4	US-09-181-585-1	Sequence 1, App1

ALIGNMENTS

RESULT 1
US-09-404-879A-72
; Sequence 72, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-72
Query Match 13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.5%; Pred. No. 1.9e-109; Indels 96; Gaps 6;
Matches 1111; Conservative 0; Mismatches 796;
1 ATGGCTCAGTTTCCACACCTTGGGTGAGCTGATGCTGGGCAATGAG 60
42 ATGGCTCAGTTTCCACACCTTGGGTGAGCTGATGCTGGGCAATGAG 101
61 GAAAGGGCCAGCATGACGAGGATGCTTACCTGAGCGATGAGGATTTACT 120
102 GAACGACTAAGCATGATTAACAGTTGATTAACCTTAACCTTACGAGGATTAACA 161
121 GGTGATCAAGCAGGAGGATTTTTCATCTGCTTACCTGAGCTGCTTACGACA 180
162 GGTGATCAAGCAGGAGGATTTTTCATCTGCTTACCTGAGCTGCTTACGACA 221
181 ATATGGGGCTTACGAGGATTAACGATGAGGATGAGGATGAGGATTTTCCATA 240
222 ATATGGGGCTTACGAGGATTAACGATGAGGATGAGGATGAGGATTTTCCATA 281
241 GCGATGAAGCTTATCAAGTGAAGTATATCAAGTCCCTCCACACTTCCCT 300
282 GGTATGAAGCTTATCAAGTGAAGTATATCAAGTCCCTCCACACTTCCCT 341
301 GGTATGAAGCTTATCAAGTGAAGTATATCAAGTCCCTCCACACTTCCCT 357
342 ATATGAAGCTTATCAAGTGAAGTATATCAAGTCCCTCCACACTTCCCT 401
358 ATATGAAGCTTATCAAGTGAAGTATATCAAGTCCCTCCACACTTCCCT 400

```

Db      402 ATGCCAATCTGTGCATTCATCAGCAATTCGCTCCAGTTGCACTATGCAACACCCCTG 461
Qy      401 -----TGGGCTCAATTCAGTTGTTGGAATGTCTCCACCTTATATCT 444
Db      462 TCTTCTGCTACTTCAGGACCAAGTATCTCCCTCATATGATGCCCTGCTCCCTATGCT 521
Qy      445 TCTGTCCTCCAGAGCAGTGCCTCCCTGGCTAACGGGGCTCCCTCCGCTCATACAGCT 504
Db      522 TCTGTAGTACATCTCTCACTTACCAATGAAATGCCAATCTCATTCAGCTTTATCCATT 581
Qy      505 CTGCTGCGTTTGGCGCATCTGACGCCACATGCGCAAAAGTTCTTCCCTGACAGATCT 564
Db      582 CTTATTTCTTCTTCAACATGCTGCTCATGATCATCTTACAGCTGATATGGAGGATTT 641
Qy      565 GGTCCAGGCTCACTTAAACATTAAGTTACAGAAAGCAATCATTTGATGTGCCACG 624
Db      642 GGTGTGTGTAATTCAGAAAGGCCCAAGTCTGTGATTTAGATTTAGATTTAGCTCACT 701
Qy      625 GCGCCCTCAG-----CAGCAGAAATGGGCTGTG 651
Db      702 TCTTCAATCTGCTTCCCTCTCAGGAACTCAGCTAAGACAGGGAAGCTCAGAGTGGCAGTT 761
Qy      652 CCTCAGTCATCAAGGCTGAAATACAGCGAGTTATTCACAGCCACGACAAACTATAGT 711
Db      762 CTTAGCCTTCAAGATTAAGATATGCGCAAAATTTATATAGCTAGACAAAGGATGAGC 821
Qy      712 GACACTTAACAGGTCCCGAGCAAACTATTTCTCATGCAATCAAGTTTACCCAGCT 771
Db      822 GGAATACCTCTCAGGTTTCAAGCTAGAAATGCCCTTCTCAGTCAAACTCTCTCAACT 881
Qy      772 CAGTGGCTTCAATATGATCTTTCTGACATTTGATCAAGATGGAATCTCAGTCAGAA 831
Db      882 CAGCTAGCTTACTATTTGACCTTGGCTACATTCATGATGAGAGCACTTGAAGCTGAA 941
Qy      832 GAATTTATCTAGCTATGACCTATATGATTTGATGTCATGTCAGCCACCTGCGGCC 891
Db      942 GAATTTATTTCTGGGATGCACTCATGACATGCGCAAGCTGACAGCCACTACATG 1001
Qy      892 GTCTGCTCTCAAGATATCATCTCTCTTCTTCAAGAAAGTTCCGCTCCGCAATGGGATG 951
Db      1002 ACGTGGCTCTCCGAGCTGTGCTCTCATCTTTCAGAGGGGAAAGCAAG----- 1050
Qy      952 TCCGTCATAGCTCTTCTTGTGATGACAGAGGCTGCGTGAAGCGCTGTCAGAGAT 1011
Db      1051 -----TTGATTTCTGTTATGAACTCTGCTTCAATATCAGAAAACACAGAA 1097
Qy      1012 GAGCAGCAGCAGAGAGAACTGCTGTGACATTTGAAGATTAAGAACCGGAGAACTTC 1071
Db      1098 GA---AGAGCTCGAAGAAACTGCAATTTCTTTGAGACAAACGAAAGCCACTAT 1154
Qy      1072 GAGCAGGCACTGTGAGCTGAGAAAGCGCGCCAGCGCTCTTGGAGCAGCAGCGCAA 1131
Db      1155 GAACGAGAGAACTATGAGACTGAGAAAGGACGCCAAGTGTGATGAGACACAGAGAGG 1214
Qy      1132 GAGCAGAGCGGTTGGCTCAGCTGAGCGCGCGGACGAGAGAGAAAGAGCGGAGGCG 1191
Db      1215 GAGGCTGAACCGAAAGCCCAAGAAAGAAAGAGAGTGGAGCGGAAACAGAGAACTG 1274
Qy      1192 CAGAGCAGAGAGCGCAAGCGCAGCTGAGCTGAGAAAGCAGCTGAGAGCAGCGGAG 1251
Db      1275 CAAGAGCAAGATGAGAAAGAGAGCTGAGAGTTGAGAAACCGTTGAGAAACAGAGAG 1334
Qy      1252 CTGAGACGGCAGCGAGAGAGAGAGAGAGAGAGATCGAGAGCGCGCGCAAAA 1311
Db      1335 CTGAGAGACAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
Qy      1312 CGGGAATCGGAAGGACAGCACTTGAATGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1371
Db      1395 CAGAGCTTGAAGAGCAAGCGCTTTAGAAATGGAGAAAGAGAGAGAGAGAGAGAGAG 1454
Qy      1372 AATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431

```

```

Db      1455 AGTCAGAGACAGAGAGAAACAAAGAGACATTGTGAGCTGAGCTCCAGAAAGAAAGTCTC 1514
Qy      1432 GAGTTTGAATGAAGCTCTGATGACAAAGAAAGATACAGCTAGAGAGAAAGCTTCAAGAT 1491
Db      1515 CACTGGAACCTGGAAGAGAGATGAAAGAAACATGAGAGATCTTCAGGACAGCTACAGAT 1574
Qy      1492 ATCAGTGTGCACTGGGACCCAGAGGCAAGAAATTTAGAGACAGAACAGTCTAGAGAG 1551
Db      1575 GTCCAAATCAGAAAGCAACACAAAGACAGAGCTGAGCTAGAAAGTTTGAATTAACAGTGTGAC 1634
Qy      1552 CTAAAGATTCGTGAATCACCACCTTACAGCAGAGAGTTGACAGAAATCTCAGCAAAATCTT 1611
Db      1635 CTGGAATTTATGAAATCAAACTTCAAGAGAGCTTAAAGAAATATCAAAATTAAGCTT 1694
Qy      1612 GGAAGACTTATTCAGAGAAACAGATCTCAGTACAGACAGTTTAAACAGTCCAGCAAGC 1671
Db      1695 ATCTATCTGCTCCCTGAGAGAGAGCTATTAACCAAGAAATTAACATGACAGCTCAGT 1754
Qy      1672 AGTTGATAGAGACTGCTTCTTACCTCAAAAGAGCCTTGGAAAGCAAGAGCTGGCC 1731
Db      1755 AACCACTGATTCAGGAGATCAGTTTACTTCTATTAAGATCATGAGAAAGAGAAATTA 1814
Qy      1732 CGGAGAGCTCCGGAGCAGCTGAGCAGAGTGGAGAGAGAGACAGCTCAAGCTGACG 1791
Db      1815 TGCCAAAGACTTAAAGAACTTATGATGCTCTGAAAGAAAGAACTGCATCTAAGCTCTCA 1874
Qy      1792 GAGTTATGTTTCAACCAAGCTGAGAGAACTGAGAGATCACTACCAACAGACAA 1851
Db      1875 GAAATGATTCATTTAAACATCAGCTAGAGAACTCAGAGAAAGCTATTAATACAGCAG 1934
Qy      1852 CTCAGAGAGAGAGTCCCTGAGAGCGCAGCTGAGAGCACTGAGAGCAAGAGCAGAGAGAG 1911
Db      1935 TTAGCCCTTGAACAACCTTCAATTAATCAAAAGTACAAATTTGAAGAAATGAAAGAAA 1994
Qy      1912 AGCTGAGTTAGAGAACAAA 1934
Db      1995 AGATTAGACCAAAAAAAAAA 2017

```

```

RESULT 2
US-09-338-933-72
; Sequence 72, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-72

Query Match      13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.5%; Pred. No. 1.9e-109;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

```

Db 162 GGTATCAAGCCGTACTTTTCTTCTAGACGAGTTCGCGGCCCGGTTTACGTGA 221
 Qy 181 ATATGGGCGCTAGCGGACATGAAATACGATGGAAGGATGATCAAGTGAATTTTCATA 240
 Db 222 AATATGGGCGCTTATCAGATCTGAAACAGATGGAGATGGACCAAGAGATGCTTATA 281
 Qy 241 GCGATGAAGCTTATCAAACTGAAAGCTACAAAGATATCAGTCCCTCCACACTTCCCTC 300
 Db 282 GCTATGAACCTATCAAGTTAAAGTTGACGGGCAACAGCTGCTGTAGTCCCTCCT 341
 Qy 301 GTCATGAACACGACACAGT---GCTATTTCCAGTGCACACAGACTTTGTATGAGAG 357
 Db 342 ATCAATGAACCAACCCCTTATGTTCTCTCACTAATCTGCTGTTGTGAGTGGAGAC 401
 Qy 358 ATTTGATGAGT---CCACACTCAAGCTGTTGCTGCTGTCGA-----400
 Db 402 ATGCCCAATCTGCTTCAATCAAGCAATGCTCCAGTTCACCTATAGCAACACCTTG 461
 Qy 401 -----TGGGCTCCATTCAGTGTGGAATGCTCCACCTTATGATCT 444
 Db 462 TCTTCTGATCTTCAAGGACAGATATTCCTCCCTAATAGATGCTGCTCCCTAGTGC 521
 Qy 445 TCTGTCCTCCACGACAGTGCCTCCCTGCTGCTAACGGGCTCTCCCTCATACAGCT 504
 Db 522 TCTGTATGACATCTCTCATTTACCAATGGAACCTGCAATGAGCTTTATCAT 581
 Qy 505 CTGCGCTGCTTGGCGATCTCTGACGACATGGCCAAAGATTTCTTCCCTTACGACATCT 564
 Db 582 CTTATTTCTTCTCAACATTTGCTCATGATCATCTTACAGCTGATGAGAGATTT 641
 Qy 565 GGTCCAGGCTCAACATTTAACTAATGTTACAGAAAGCAAACTCATTCATGTCGACAG 624
 Db 642 GGTGGTGTAGATCCAGAAAGGCCAGTCTGTGATTTAGATCTAGTACCTCACT 701
 Qy 625 GCCCTCCAG-----CAGCAGATGGGCTGTG 651
 Db 702 TCCCTCAATGCTTCCCTCTCAGGGAATCACTAAGACAGGACCTCAGAGTGGAGAT 761
 Qy 652 CCTCATGCTCAAGGCTGAATACAGGCAATTAACAGACGACAAACTATGAT 711
 Db 762 CCTCAGCCTTCAAGATTAAGTACGCAAAATTTAATGCTTACAAAGGACATGAC 821
 Qy 712 GGAACCTTAAAGGCTCCCGACGACAACTATTCATGCAATCAAGTTTACCCAGCT 771
 Db 822 GGAATCTCTCAGGTTTCAAGCTAAGATGCCCTTCTTCAATCTCTCTCAACT 881
 Qy 772 CAGCTGCTTCAATATGAACTTTCTGACATTTGATCAAGATGAAACTACTGACAA 831
 Db 882 CAGCTAGCTACTATTTGAGCTGTGCTGACATGATGTGACGACAGTTGAAGCTGA 941
 Qy 832 GAATTTATCTAGCTATGACCTAATTTGATGTTGCTGATGCTGACGACCTGCCGCC 891
 Db 942 GAATTTATTTCTGCGATGACCTCACTGACATGCGCAAAAGCTGACAGCACTACAC 1001
 Qy 892 GTCCTGCTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTCGCTCCGCAAGTGGAG 951
 Db 1002 AGCTTCCCTCCGAGCTTGTCTCTCATTTTCAAGGGGGAAGAG-----1050
 Qy 952 TCCGCTATAGCTCTTCTTCTGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGAT 1011
 Db 1051 -----TTGATTTCTGTTATGAACTCTGCTTCAATCAGAAAACAGAA 1097
 Qy 1012 GAGCAGACCCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGACGGAGAACTTC 1071
 Db 1098 GA--AGAGCTCAGAAAGAACTGCAATTTTGAAGACAAACGAAAGCCCACTAT 1154
 Qy 1072 GAGCGAGGAGTGTGAGCTGAGAGAGCGCCGACAGCCCTTGTGAGCAGCGCAAA 1131
 Db 1155 GAACGAGAAACATGAGCTGAGAGAGCAAGCCAGATGTTATGAGACAGCAGAGAG 1214
 Qy 1132 GAGCAGAGCGGTTGCTCAGCTGAGCGCGCCGACAGAGAGAAAGCGGAGCGC 1191
 Db 1215 GAGGCTGAACGCAAGCCAGAAAGAGAAAGAGTGGAGCGGAAACAGAGAAATCTG 1274

Qy 1192 CAGAGCAGAGAGCCAGCGGAGCTGAGCTGAGAGAGCAGCTGAGAGAGCGGGG 1251
 Db 1275 CAGAGCAGAGAGTGAAGAGAGCTGAGGTTGAGAAACGTTGAGAAACAGAGAG 1334
 Qy 1252 CTGAGCGGACGAGAGAGAGAGAGAGAGATGAGAGCGGAGCGGCAAAA 1311
 Db 1335 CTGAGAGAGCAGCGGAGAGAGAGAGAGATGAGAAAGATGAGAAAGCAGAGCAG 1394
 Qy 1312 CCGGAACTGGAAGGACGACACTTGAATGGAAACGAAACCGAGAGACGAACTCTG 1371
 Db 1395 CAGGAGCTTGAAGACAAACGCGTTTGAATGGGAAAGATCTCCGTGAGAGAGCTG 1454
 Qy 1372 AATCAGAGAAACAGAGAGAGAGAGCGCCGCTGAGAGGCAAGAGAGAGCTG 1431
 Db 1455 ACTCAGAAAGCCAGGAAACAGAGACATTTGACAGCTGAGCTCCAGAAAGAAAGCT 1514
 Qy 1432 GAGTTGAGTTGAGAGCTGATGATGACAAAGATCAGCTGAGAGAGAACTTCA 1491
 Db 1515 CACCTGGAAGCTGGAAGCAGTGAATGGAAGAAATCAGCAGATCTCAGGAGACTTCA 1574
 Qy 1492 ATCAGGTGCACTGGCAACCAAGGACAGAAATGAGAGCAGACAGATCTAGAG 1551
 Db 1575 GTCCAAATCAGAAACCAACAAAGACAGTGAAGATTTGATTAACAGTGTGAC 1634
 Qy 1552 CTAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1611
 Db 1635 CTGAAATTTAGGAATCAAACTTCAACAGAGCTTAAAGAAATCAAAATTAAGCT 1694
 Qy 1612 GGAAGACTTATTCAGAGAAACAGATGATGATGATGATGATGATGATGATGATGAT 1671
 Db 1695 ATCTATCTGTGCTCCGAGAGAGAGATTTAAACGAAAGATTTAAACATGAGCTCAG 1754
 Qy 1672 AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1731
 Db 1755 AACACCTGATTCAGGATCAGTTACTTAAAGATCAAGAAAGAGAAATTA 1814
 Qy 1732 CGGACAGAGCTCCGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
 Db 1815 TGCCAAAGCTTAAAGAACTTATGATGATGATGATGATGATGATGATGATGATGAT 1874
 Qy 1792 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1851
 Db 1875 GAATGAGATTTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1934
 Qy 1852 CTCAGAAAGCAGAGGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1911
 Db 1935 TTAGCCCTTGAACAACTTCAATTAACAGTGAACAAATGAGAGAAATGAGAAAG 1994
 Qy 1912 AGCTGAGTTGAG 1934
 Db 1995 AGATTGAGCAAAAAA 2017

RESULT 3
 US-09-215-681-72
 ; Sequence 72, Application US/09215681A
 ; Patent No. 6528253
 ; GENERAL INFORMATION:
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Fundakis, Tony N.
 ; APPLICANT: King, Gordon E.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
 ; TITLE OF INVENTION: OF OVARIAN CANCER
 ; FILE REFERENCE: 210121.463
 ; CURRENT APPLICATION NUMBER: US/09/215,681A
 ; CURRENT FILING DATE: 1998-12-17
 ; NUMBER OF SEQ ID NOS: 310
 ; SOFTWARE: FaSTSeq for Windows Version 3.0
 ; SEQ ID NO 72
 ; LENGTH: 2017
 ; TYPE: DNA
 ; ORGANISM: Homo sapien

US-09-215-681-72

Query Match 13.1%; Score 477.4; DB 4; Length 2017;
Best Local Similarity 55.5%; Pred. No. 1.9e-109;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

QY 1 ATGGCTAGTTTCCACACCTTTGGGTAGCTGATGTCTGGGCCATACTGTGGAG 60
DB 42 ATGGCTAGTTTCCACACCTTTGGGTAGCTGATGTCTGGGCCATACTGTGGAG 101
QY 61 GAAAGGCGCAAGCATGACAGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTA 120
DB 102 GAACCTACTAAGCATGATTAACAGTTTGATTAACCTCAACCTTCAGAGGTTTACATA 161
QY 121 GGTATCAAGGAGAACCTTTTTCATCTGGGTACTCTCAGCCCTGTCTTACACAA 180
DB 162 GGTATCAAGGCGCGGACTTTTTCCTACAGTCAAGTCTGGCGCCCGGTTTATGCTGAA 221
QY 181 ATATGGGCGCTAGCGGACATGATTAAGATGAGATGATCAAGTGAATTTTCCATA 240
DB 222 ATATGGGCGCTTATCATGATCTGAACAAGATGAGAGATGACACAGAACGTTCTTATA 281
QY 241 GCCATGAAGCTTATCAAACTGAAGCTTACAGATATCAGCTCCCTCCACACTTCCCT 300
DB 282 GCTATGAAGCTTATCAAGTAAAGTTGACAGGCGCAACAGCTGCTGTAGTCTCCCTCT 341
QY 301 GTCATGAAGCAAGCAACAGT---GCTATTTCCAGTGCACAGCAATTTGGTATAGAGG 357
DB 342 ATCATGAAGCAACACCCCTATGTTCTCTCACTAACTCTGCTGTTTGGAGATGGGAAGC 401
QY 358 ATTTGCTAGCATG-----CCACCACTCAGCTGTTGCTCTGCTCCAA----- 400
DB 402 ATGCCCATCTGTCCATTCATCAGCATTTGCTCCAGTTGCACTATAGCAACACCTTG 461
QY 401 -----TGGGCTCATTCCAGTTGTGGAATGTCTCAACCTTATGATCT 444
DB 462 TCTTCTGCTACTTTCAGGACCAAGTATCTCTCCCTAATGATGCTGCTCCCTAGTGCT 521
QY 445 TCTGTCCTTCCAGCAGAGTGTCTCCCTGGCTAACGGGGCTCCCTCCCTATACAGCT 504
DB 522 TCTGTTAGTACATCTCATTTACCAATGGAATGCACTGCTCATTCACCTTTATCCATT 581
QY 505 CTGCTGCGTTTGGSCATCTGSCAGCCACATGGCCAAAGATTTCTCTTACAGCATCT 564
DB 582 CTTTATTTCTTTCACCAATGCTCTATGATCTTACAGCTGATATGAGAGATTT 641
QY 565 GGTCCAGGATCAAACTTAACACTAAGTTTACAGAGGCAACATCTGATGTGCCAGC 624
DB 642 GGTGTGCTAGTATCCAGAGGCGCAAGTCTGATTGATTAGATCTAGTAGTCAACT 701
QY 625 GGCCTCCAG-----CAGCAGATGGGCTGTG 651
DB 702 TCTCTAAGCTGCTTCCCTCTCAGGGAACCTACCTAAGACAGGACCTCAGAGTGGCAGTT 761
QY 652 CTTGATCATCAAGGCTGAATATACAGGAGTATTTCAACAGCCAGCAAAACTATAGT 711
DB 762 CTTGAGCTTCAAGATTAAGATGCGCAAAATTTAATAGTTCAGCAAGAGCATGAGC 821
QY 712 GGAACATTAAACAGTCCCGAGGCAAGAACTATTTCTATGCAATCAAGTTTACCCAGGCT 771
DB 822 GGAATACCTCTCAGGTTTTCAGCTGAGTAAAGCTTCTTCAAGTCAAACTCTCAAACT 881
QY 772 CAGCTGCGCTTCAATATGGAATTTTCTGACATGATCAAGATGGAATACTACTGCAAA 831
DB 882 CAGCTAGCTACTATATTTGAGCTCTGCTGATCATGATGAGTGAAGAGTGAAGCTTAA 941
QY 832 GAAATTATCTAGCTATGACCTAATATGATGATGATGATGATGATGATGATGATGATG 891
DB 942 GAAATTATCTAGCTATGACCTAATATGATGATGATGATGATGATGATGATGATGATG 1001
QY 892 GTCTGCTCTCCAGAAATACATCCCTCTCTCTTCCAGAAAGATTCGCTCCGCGCAGTGG 951
DB 1002 AGCTTGCCTCCCGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050

QY 952 TCCGTATTAAGCTCTTCTCTGTGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGAT 1011
DB 1051 -----TTGATTTCTGTTAATGAACTCTGCTTCTATATGCAAGAAACACAGAA 1097
QY 1012 GAGCAGCAGCAGAGAAAGAACTGCTGTGACATTTGAAATGAAGAGCGGGAACCTTC 1071
DB 1098 GA---AAGCCTCAGAAAGAACTGCCAGTTACTTTTGAAGCAAAACGGAAGCCACTAT 1154
QY 1072 GAGGAGGCGAGTGTGAGCTGAGAAAGCCCGCAAGCCTCTTTGAGCGACGCGCAA 1131
DB 1155 GAAGAGAGAAACATGAGCTGAGAAAGCCAGCCAACTGTTGATGAGCAGCAGCAGAG 1214
QY 1132 GAGCAGGAGCGGTTGGCTGAGCTGAGCGCGCCAGCAGAGAGAGAAAGCGGAGCGC 1191
DB 1215 GAGGCTAAGCAGAAAGCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
QY 1192 CAGGAGCAGAGAGCGCAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251
DB 1275 CAGGAGCAGAGATGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
QY 1252 CTGAGCGGAGCGAG 1311
DB 1335 CTGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
QY 1312 CCGGAACTGGAAGGAG 1371
DB 1395 CAGGAGCTTGAAG 1454
QY 1372 AATCAG 1431
DB 1455 AGTCAG 1514
QY 1432 GAGTTGAGTTAAGAGCTCTGAATGACAAAGATAGCTAGAGAGAGAGAGAGAGAG 1491
DB 1515 CACCTGAG 1574
QY 1492 ATCAGTGTGAG 1551
DB 1575 GTCCAAATCAGAAAG 1634
QY 1552 CTAAAGATGCTGAATACCCACTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
DB 1635 CTGGAATTTATGGAATCAACCACTTCAACAGAGCTTAAAGAAATTAAGCTT 1694
QY 1612 GGAAGATTTATTCAG 1671
DB 1695 ATCTATCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
QY 1672 AGTTGATAG 1731
DB 1755 AACACCTGATTCAGAGATCAGTTTACTTATTAAGATCATCAGAAAGAGAGAGAT 1814
QY 1732 CCGCAGAGAGCTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
DB 1815 TGCCAAAGAGCTTAAAGAACTATGATGCTCTTGAAGAAAGAGAGAGAGAGAGAG 1874
QY 1792 GAGATTGATTTTCAACCAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
DB 1875 GAAATGATTTATTAACATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
QY 1852 CTCAG 1911
DB 1935 TTAGCCCTTGAACAATTCATTAATCAAAAGTGAACAAATTTGAAGAAATTCAGAA 1994
QY 1912 AGCTGAGATTAG 1934
DB 1995 AGATTAG 2017

RESULT 4
us-08-630-915A-193
; Sequence 193, Application us/08630915A

SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA: US/08/630,915A
 APPLICATION NUMBER: US/08/630,915A
 FILING DATE: 03-APR-1996
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Mirock, S. Leslie
 REFERENCE NUMBER: 18,872
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 747 bases
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-630-915A-39

Query Match 7.1%; Score 258; DB 4; Length 747;
 Best Local Similarity 63.9%; Pred. No. 7.6e-55;
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;
 QY 2989 GGAGAGGATTATTCGATGACATACAGAGTTCTGAGCAGAGATTAACTTT 3048
 DB 115 GGAGAGGATTATTCGATGACATACAGAGTTCTGAGCAGAGATTAACTTT 174
 QY 3049 CAGAGAGGAGATGATGTTGGTTACGAGAGAGATGAGCTGAGCAGAGAGGCT 3108
 DB 175 ACAGAGGATGAGAGATTTGGTACCCAGAGAGAGAGAGAGAGAGAGATTT 234
 QY 3109 GCGCAGAGATCGAGATCTTCCCTTCTAATGAGAGCTTAAAGATTCAAGGCTCT 3168
 DB 235 GGAGATGAGAGATTTTTCATCAATATGTCAGAGAGAGATCAAGAGATTTT 294
 QY 3169 GGAATGCTGGAG 3228
 DB 295 GGAATGCTGGAG 354
 QY 3229 TCCTAGCGTCTACTGTCGTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3288
 DB 355 GCATATGTTGCTTCTGTTCTGACCACTTACCTTGCAGAGAGAGAGAGAGAG 414
 QY 3289 CCGAG 3348
 DB 415 CTAAAG 474
 QY 3349 CAGATAGGAGTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3408
 DB 475 CAGAGAGAGATGTTTCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 534
 QY 3409 ACCCAAGTGAAGTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3468
 DB 535 AACCTGCTTTCACTCT-----GTAATGACAGATTTCTATG 573
 QY 3469 TACGATTAACCCAG 3528
 DB 574 TATGATTAACAG 633
 QY 3529 GTCTTCAACAG 3588
 DB 634 GTTATGAACAG 693
 QY 3589 TTCCATCAATTAATGTAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3642
 DB 694 TTCTCTCAACTAG 747

RESULT 6
 US-09-404-879A-60

Sequence 60, Application US/09404879A
 Patent No. 648846
 GENERAL INFORMATION:
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: King, Gordon E.
 APPLICANT: Algate, Paul A.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
 FILE REFERENCE: 210121.462C1
 CURRENT APPLICATION NUMBER: US/09/404,879A
 CURRENT FILING DATE: 1999-09-24
 NUMBER OF SEQ ID NOS: 393
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 60
 LENGTH: 480
 TYPE: DNA
 ORGANISM: Homo sapien
 US-09-404-879A-60

Query Match 4.0%; Score 144; DB 4; Length 480;
 Best Local Similarity 62.1%; Pred. No. 1.9e-26;
 Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;
 QY 1 ATGGCTAGATTCCACACCTTTCGATGAGCTGATGATGAGGAGATTAAGTGGAG 60
 DB 14 ATGGCTAGATTCCACACGAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 73
 QY 61 GAAAGGCGCAAGATGACAGAGATTCCTTACCTGAGAGCCGAGATTTATTAAT 120
 DB 74 GAAAGGCGCAAGATGACAGAGATTCCTTACCTGAGAGCCGAGATTTATTAAT 133
 QY 121 GGTATCAAGAGAGAGATTTTTCCTCAATCTGAGGTTTACCTGAGGCTGTTAGCA 180
 DB 134 GGTATCAAGAGAGAGATTTTTCCTCAATCTGAGGTTTACCTGAGGCTGTTAGCA 193
 QY 181 ATATGGCGCTAGGAGAGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
 DB 194 ATATGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253
 QY 241 GCTATGAACTCATCAAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 DB 254 GCTATGAACTCATCAAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 313
 QY 301 GCTATGAACTCATCAAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 DB 314 ATATGAACTCATCAAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 367
 QY 361 GCTATGAACTCATCAAGATTAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 398
 DB 368 GAAAGCAGTCCCATCTGTCATTCATTCATTCATTCATTCATTCATTCATTCAT 405

RESULT 7
 US-09-338-933-60
 Sequence 60, Application US/09338933
 Patent No. 6488931
 GENERAL INFORMATION:
 APPLICANT: Mitcham, Jennifer Lynn
 APPLICANT: King, Gordon E.
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
 TITLE OF INVENTION: OVARIAN CANCER
 FILE REFERENCE: 210121.462C1
 CURRENT APPLICATION NUMBER: US/09/338,933
 CURRENT FILING DATE: 1999-06-23
 NUMBER OF SEQ ID NOS: 312
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 60
 LENGTH: 480
 TYPE: DNA
 ORGANISM: Homo sapien
 US-09-338-933-60

Query Match 4.0%; Score 144; DB 4; Length 480;

Best Local Similarity 62.1%; Pred. No. 1.9e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
QY 1 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 60
  |||
Db 14 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 73
QY 61 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
  |||
Db 74 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 180
  |||
Db 134 GGTGATCAAGCAGCCTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 193
QY 181 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 240
  |||
Db 194 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 253
QY 241 GCCATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 300
  |||
Db 254 GCTATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 313
QY 301 GTCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGGAT 360
  |||
Db 314 ATCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGG 367
QY 361 GCTAGATGCCACACTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
  |||
Db 368 GGAAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 405
```

RESULT 8

US-09-215-681-60
; Sequence 60, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-215-681-60

Query Match 4.0%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 1.9e-26;

Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
QY 1 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 60
  |||
Db 14 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 73
QY 61 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
  |||
Db 74 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 133
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 180
  |||
Db 134 GGTGATCAAGCAGCCTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 193
QY 181 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 240
  |||
Db 194 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 253
```

```
QY 241 GCCATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 300
  |||
Db 254 GCTATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 313
QY 301 GTCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGGAT 360
  |||
Db 314 ATCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGG 367
QY 361 GCTAGATGCCACACTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
  |||
Db 368 GGAAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 405
```

RESULT 9

US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;

Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
QY 1 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 60
  |||
Db 42 ATGGCTCAAGTTCCCAACCTTTGGTGTGATGATGTCGGGCCATTAAGTGGAG 101
QY 61 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
  |||
Db 102 GAAAGGGCCAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 161
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 180
  |||
Db 162 GGTGATCAAGCAGCCTTTTTCATCTGGGTACCTCAGCTGTCTAGACAA 221
QY 181 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 240
  |||
Db 222 ATATGGGCGCTATGCGGACATGATGATGATGATGATGATGATGATGATGAT 281
QY 241 GCCATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 300
  |||
Db 282 GCTATGAAGCTTATCAAGTGAAGTACAGATATCAGTCCCTCCACACTTCCCT 341
QY 301 GTCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGGAT 360
  |||
Db 342 ATCATGAAGCAGCAACAGTGTGATTTCCAGTGCACAGCAATTTGATAGAGG 395
QY 361 GCTAGATGCCACACTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
  |||
Db 396 GGAAGCATGCCCAATCTGTTCATTCATCAGCAATGCC 433
```

RESULT 10

US-09-338-933-5
; Sequence 5, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn

APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
TITLE OF INVENTION: OVARIAN CANCER
FILE REFERENCE: 210121.462C1
CURRENT APPLICATION NUMBER: US/09/338,933
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 312
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO: 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-338-933-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```
Qy 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTGTGAGCCATTAATCTGTGAG 60
    |||
Db 42 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTGTGAGCCATTAATCTGTGAG 101
    |||
Qy 61 GAAAGGCCAAGCATGACAGAGCTTCTTAAAGCCGATAGCGGATTTACT 120
    |||
Db 102 GAAAGGCTAAGCATGATTAACAGTTTAACTCAAGCTTCAAGAGGTTACATA 161
    |||
Qy 121 GGTATCAAGGAGAACTTTTTCCTCAATCTGGGTACTCAGCTCTTACGACAA 180
    |||
Db 162 GGTATCAAGCCGCTACTTTTCTTCAAGTCAAGTCTGCGGCGCTTTTACGAA 221
    |||
Qy 181 ATATGGCGCTAGCGGACATGATTAACATGAGATGATCAAGTGAATTTCCATA 240
    |||
Db 222 ATATGGCGCTTATCATGATCTGAACAGATGAGAGATGACAGCAAGATCTTATA 281
    |||
Qy 241 GCCATGAAGCTTATCAACTGAAGCTACAGATATCAAGTCTCCCTCCACCTTCCCT 300
    |||
Db 282 GCTATGAAGCTATCATGATTAAGTTGAGGCGCAACAGCTGCTAGTCTCCCTCCT 341
    |||
Qy 301 GTCATGAAGCAAGCAAGCTGCTTATTTCCAGTGCACAGATTTGATAGAGGAT 360
    |||
Db 342 ATCATGAAGCAACACCC-----CTATGTTCTCTCCACTAATCTGCTGTTTGGGATG 395
    |||
Qy 361 GCTAGCATGCCACCACTCAAGCTGTGCTCTGCGC 398
    |||
Db 396 GGAAGCATGCCCAATCTGTCAATTCATCAGCCATTGCC 433
    |||
```

RESULT 11
US-09-215-681-5
Sequence 5, Application US/09215681A
Patent No. 6528253
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Frudakis, Tony N.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF OVARIAN CANCER
FILE REFERENCE: 210121.463
CURRENT APPLICATION NUMBER: US/09/215,681A
CURRENT FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 310
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO: 5
LENGTH: 531
TYPE: DNA
ORGANISM: Homo sapien
US-09-215-681-5

Query Match 4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;
Qy 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTGTGAGCCATTAATCTGTGAG 60

```
Db 42 ATGGCTCAGTTTCCACAGGATGAATGAGGGCAAAATATGTGGCTATTAATCTGAA 101
    |||
Qy 61 GAAAGGCCAAGCATGACAGAGCTTCTTAAAGCCGATAGCGGATTTACT 120
    |||
Db 102 GAAAGGCTAAGCATGATTAACAGTTTAACTCAAGCTTCAAGAGGTTTACATA 161
    |||
Qy 121 GGTATCAAGGAGAACTTTTTCCTCAATCTGGGTACTCAGCTCTTACGACAA 180
    |||
Db 162 GGTATCAAGCCGCTACTTTTCTTCAAGTCAAGTCTGCGGCGCTTTTACGAA 221
    |||
Qy 181 ATATGGCGCTAGCGGACATGATTAACATGAGATGATCAAGTGAATTTCCATA 240
    |||
Db 222 ATATGGCGCTTATCATGATCTGAACAGATGAGAGATGACAGCAAGATCTTATA 281
    |||
Qy 241 GCCATGAAGCTTATCAACTGAAGCTACAGATATTAAGTCTCCCTCAACTTCCCT 300
    |||
Db 282 GCTATGAAGCTATCATGATTAAGTTGAGGCGCAACAGCTGCTGATGCTCCCTCCT 341
    |||
Qy 301 GTCATGAAGCAAGCAAGCTGCTATTTCCAGTGCACAGATTTGATAGAGGAT 360
    |||
Db 342 ATCATGAAGCAACACCC-----CTATGTTCTCTCCACTAATCTGCTGTTTGGGATG 395
    |||
Qy 361 GCTAGCATGCCACCACTCAAGCTGTGCTCTGCGC 398
    |||
Db 396 GGAAGCATGCCCAATCTGTCAATTCATCAGCCATTGCC 433
    |||
```

RESULT 12
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS: 21
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 4.0%; Score 144; DB 2; Length 3489;
Best Local Similarity 45.1%; Pred. No. 5e-26;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

QY 975 GATCAGAGGCTTCTGAGAGAGCCGCTGCTCAGAGATGAGCAGACCAAGAGAAACT 1034
DB 1578 GCGAGGACCAACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1637
QY 1035 GCCTGTGACATTTGAGATTAAGAGCGGAGAACTTTCAGCGAGCGAGTGTGAGCTGGA 1094
DB 1638 GCGAGGACCAACAGCAGAGAGCCAGCAGAGCGGAGAGCCAGCAGAGAGCCAGCA 1697
QY 1095 GAAAGCCGCCCAAGCCCTCTTGAGAGCAGCAGCAAGAGAGAGAGAGAGAGCT 1154
DB 1698 GCGGAGACCAACAGCAGAGAGCCAGCAGAGCGGAGAGCCAGCAGAGAGCCAGCA 1757
QY 1155 GAGGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214
DB 1758 GCGGAGACCAACAGCAGAGAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1817
QY 1215 GCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
DB 1818 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1877
QY 1275 GAGGAGAGAGAGAGATGAGAGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAG 1334
DB 1878 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1937
QY 1335 ACTGTAATGGGAAACGAGACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
DB 1938 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1997
QY 1395 GGGCAGCCTGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
DB 1998 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2057
QY 1455 TGAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
DB 2058 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2117
QY 1515 GAGGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
DB 2118 GCGAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2177
QY 1575 CTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
DB 2178 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2237
QY 1635 GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
DB 2238 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2297
QY 1695 TACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
DB 2298 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2357
QY 1755 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
DB 2358 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2417
QY 1815 GCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1874
DB 2418 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2477
QY 1875 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
DB 2478 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2537
QY 1935 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994

```

```

DB 2538 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2597
QY 1995 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2054
DB 2598 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2657
QY 2055 TGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2114
DB 2658 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2717
QY 2115 GCTTTCCATCCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
DB 2718 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2761

```

RESULT 13

```

US-09-298-568-1
; Sequence 1, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieft, Elliott D.
; APPLICANT: Bailestae, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; EARLIER FILING DATE: 1999-04-21
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kapoli's sarcoma-associated herpesvirus
US-09-298-568-1

```

Query Match 4.0%; Score 144; DB 4; Length 3489;
Best Local Similarity 45.1%; Pred. No. 5e-26;
Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

```

QY 975 GATCAGAGGCTTCTGAGAGAGCCGCTGCTCAGAGATGAGCAGACCAAGAGAAACT 1034
DB 1578 GCGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCAGAGAGCCACAGCA 1637
QY 1035 GCCTGTGACATTTGAGATTAAGAGAGCGGAGAACTTTCAGCGAGCGAGTGTGAGCTGGA 1094
DB 1638 GCGAGAGCCACAGCAGAGAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1697
QY 1095 GAAAGCCGCCCAAGCCCTCTTGAGAGCAGCAGCAAGAGAGAGAGAGAGAGAGAGAG 1154
DB 1698 GCGGAGACCAACAGCAGAGAGCCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1757
QY 1155 GAGGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214
DB 1758 GCGGAGACCAACAGCAGAGAGCCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCA 1817
QY 1215 GCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
DB 1818 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1877
QY 1275 GAGGAGAGAGAGATGAGAGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
DB 1878 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1937
QY 1335 ACTTGAATGGGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1394
DB 1938 GCGAGATGAGCAGCAGAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1997
QY 1935 GGGCAGCCTGTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454

```

Db 1998 GCAGGATGAGCAGCAGCAGATGAGCAGAGGATGAGCAGCAGCAGATGAGCAGCA 2057
 Qy 1455 TGACAAAAGCATTCAGCTTAGAAGAAAATTTCAGGATATCAGGTTCGATCGGACACCA 1514
 Db 2058 GCAGATGATGACGACGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2117
 Qy 1515 GAGCAGAAATTGAGCAGCAGCAAGCTCTAGAGCTTAAATTTGCTGAATCACCACCA 1574
 Db 2118 GAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCA 2177
 Qy 1575 CTTAAGCAGCAGCTTGCAGGAATCTCAGCAAACTGCTTGAAGATTTATTCAGAGAAA 1634
 Db 2178 GATATGACAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGA 2237
 Qy 1635 GATATCTAGTACCATCTTAAACAGCTCAGCAGAACTTTGCTTAAGACTCGCTTCT 1694
 Db 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2297
 Qy 1695 TACCTCAAAAAGCCTTGAAGCAAAAGAGCTGCGCCGACAGCTCCGGAGCAGCT 1754
 Db 2298 GCAGAGATTTAGAGCAGCAGCAGCAGATGAGATCAGAGCAGCAGATTTAGAGAGCA 2357
 Qy 1755 GCAGCAGCTGAGAGAGAGACCAAGCTCAAGCTGAGAGATTTATTTCAACACCA 1814
 Db 2358 GCAGCAGAGTTAGAGAGCAGCAGCAGCAGAGATTTAGAGAGCAGCAGATTTAGAGCA 2417
 Qy 1815 GCTGAAGAACTGAGAGAGATCATAGCAACAGCACTCCGAAGCAGAGTCCCTTGA 1874
 Db 2418 GCAGAGCAGCAGATTTAGAGAGCAGCAGCAGCAGATTTAGAGAGCAGCAGAGTTTGA 2477
 Qy 1875 GGCAGCGGACTGAAGCAGAAAGAGCAGAGAGAGAGCTGAGATTTAGAGAAACAAA 1934
 Db 2478 GCAGCAGAGCAGCAGATTTAGAGAGCAGCAGCAGCAGATTTAGAGAGAGTGAAGA 2537
 Qy 1935 GGAAGACCTCTAGAGACAGATTCAGAGAAAGGACCAAGCAATGCTGAGCATGTGACCA 1994
 Db 2538 GCAAGAGCAGAGGTGAGAAAGCAGAGCAGAGCAGAGCAAGCAATTTAGAGAGGT 2597
 Qy 1995 GAGAGAGCAGCAGCCCCCGGAAACCCCAAGAGAGCAGACAGATCTGAAGAGGAAAGCAG 2054
 Db 2598 GAGAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAGATTTAGAGAGGTGAAGA 2657
 Qy 2055 TGTCAAGAAAGAGGCGGAGAGAGAGAGCCAAAGCCGGAATGCAAGCAAGCAGAGTCG 2114
 Db 2658 GCAGAGAAAGCAGAGATTTAGAGAGGTGAGAGAGCAGAGAGCAGAGATTTAGAGAGGT 2717
 Qy 2115 GCTTTTCATCCGATCAGAGACCACTAAGCTGCGCACCCAGG 2158
 Db 2718 GGAAGAGCAGAGCAGCAGAGATTTAGAGAGGTGAGAGAGCAGG 2761

RESULT 14
 US-09-410-399-1
 ; Sequence 1, Application US/09410399
 ; Patent No. 6482587
 ; GENERAL INFORMATION:

; APPLICANT: Robertson, Eyle S.
 ; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
 ; FILE REFERENCE: UM-03778
 ; CURRENT APPLICATION NUMBER: US/09/410,399
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 3489
 ; TYPE: DNA
 ; ORGANISM: Kapoisi's sarcoma-associated herpesvirus
 US-09-410-399-1

Query Match 4.0%; Score 144; DB 4; Length 3489;
 Best Local Similarity 45.1%; Pred. No. 5e-26;

Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;
 Qy 975 GATTCAGAGGCTGCTTGAAGAGCCGTGTGATGAGGATGAGCAGAGCAGAGAAACT 1034
 Db 1578 GCAGAGCCCAAGCAGAGAGAGCCCAAGCAGAGAGCCCAAGCAGAGAGAGCCACAGCA 1637
 Qy 1035 GCTGTGACATTTGAAATGAAGAGCGGAGAACTTTGAGCGAGGCACTGTGAGCTTGA 1094
 Db 1638 GCAGAGACCCACAGCAGCGGAGAGCCACAGCAGCGGAGACCCACAGCAGCGGAGCCACAGCA 1697
 Qy 1095 GAAGCGCCGCCAGCGCTCTTGAAGCAGACCGCAAGAGCAGAGAGCGGTGCTCAGCT 1154
 Db 1698 GCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757
 Qy 1155 GGAACGGCGGAGAGAGAGAAAGCGGAGCGCCAGAGCAGAGAGCCAGCGGCA 1214
 Db 1758 GCGGAGCCCAAGCAGCGGAGCCACAGCAGAGATGAGCAGCAGAGAGAGTGAAGCA 1817
 Qy 1215 GCTGAGCTGAGAGAGAGCTGAGAGAGAGCGGAGCTGAGCGGAGCAGCAGAGAGAG 1274
 Db 1818 GCAGGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 1877
 Qy 1275 GAGGAGAAAGAGATCGAGAGCGCGAGCCGCAAAACGGGAACTGAGAAAGCAGCA 1334
 Db 1878 GCAGGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 1937
 Qy 1335 ACTTGAATGGAAGAGAAACCGGAGACAGAACTCTGATATCAGAGAAACAGAGAGCA 1394
 Db 1938 GCAGGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 1997
 Qy 1395 GGGCAGCTGCTTGAAGCAGAGAGAGAGACTCTGAGATTTAGTTAGAGCTTGA 1454
 Db 1998 GCAGGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 2057
 Qy 1455 TGACAAAAGCATCTAGAGAGAACTTCAAGATATCAGGTGTCAGTCCACCA 1514
 Db 2058 GCAGGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 2117
 Qy 1515 GAGCAGAAATTTAGAGCAGCAGAACAGTCTAGAGACTTAAGATTCTGTAATCACCCA 1574
 Db 2118 GAGCAGCAGATGAGCAGCAGAGCAGCAGAGTGAACACAGCAGAGATGAGCAGCAGCA 2177
 Qy 1575 CTTACAGCAGAGTTGAGAGATCTCAGCAAACTCTTGAAGATTTTCCAGAGAAACA 1634
 Db 2178 GAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGATGAGCAGCAGAGATGA 2237
 Qy 1635 GATATCTAGTACCATCTTAAACAGCTCAGCAGAACTTTGCTTAAGACTCGCTTCT 1694
 Db 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2297
 Qy 1695 TACCTCAAAAAGCCTTGAAGCAAAAGAGCTGCGCCGACAGCTCCGGAGCAGCT 1754
 Db 2298 GCAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGATCAGAGAGATTTAGAGAGCA 2357
 Qy 1755 GAGCAGGTTGAGAGAGACCAAGCTCAAGCTCAGAGATTTGTTTCAACAAACA 1814
 Db 2358 GAGCAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGATTTAGAGAGCA 2417
 Qy 1815 GCTGAAGAACTGAGAGAGATCATAGCAACAGCACTCCGAAGCAGAGTCCCTTGA 1874
 Db 2418 GCAGAGAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGATTTAGAGCA 2477
 Qy 1875 GGCAGCGGACTGAAGCAGAAAGCAGAGAGAGAGAGCTTGAAGTTAGAGAGCAAAA 1934
 Db 2478 GAGCAGAGAGCAGAGATTTAGAGAGCAGAGCAGAGATTTAGAGAGCAGAGGTGAAGA 2537
 Qy 1935 GGAAGACCTCTAGAGAGAGATTTAGAGAAAGGAGCAAGCAATGCTGAGCATGTGACCA 1994
 Db 2538 GCAAGAGCAGAGGTGAGAGAGCAGAGCAGAGCAGAGCAGAGATTTAGAGAGGT 2597
 Qy 1995 GAGAGAGCAGCAGCCCCCGGAAACCCCAAGAGAGCAGACAGACTGAAGAGGAAAGCAG 2054
 Db 2598 GAGAGAGCAGAGCAGAGAGCAGAGAGCAGAGAGCAGAGATTTAGAGAGGTGAAGA 2657

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:03:53 / Search time 40 Seconds
(without alignments)
4813.383 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQFPTFGSLDVAIVTIVE.....VGLPSPNVYKLTDTMDPSQO 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq.190u03:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6258.5	99.8	1214	21	AAV57444
2	6258.5	99.8	1715	21	AAV57449
3	5860	93.5	1215	20	AAV32156
4	5847.5	93.3	1220	20	AAV32155
5	5442.5	86.8	1144	20	AAV32154
6	4277	68.2	1031	22	AAU87168
7	3097.5	49.4	1035	22	AAW43519
8	3097.5	49.3	932	22	ABG20575
9	3091.5	49.3	1681	23	AAO17881

10	3070.5	49.0	648	20	AAV32157	Human SH3D1A prote
11	3068	48.9	1197	22	AAW79199	Human protein SEQ
12	3049.5	48.6	1197	21	AAV57445	Mouse Ees2l protei
13	3049.5	48.6	1658	21	AAV57450	Mouse Ees2l protei
14	3046	48.6	1683	21	AAV71160	Rat phosphodiester
15	3034.5	48.4	642	22	AAW93229	Human polypeptide,
16	3012	48.0	641	20	AAV32158	Human SH3D1A prote
17	2812	44.9	595	22	AAW94691	Human protein sequ
18	2273.5	36.3	471	22	AAW93675	Human polypeptide,
19	2058.5	32.8	462	17	AAW05395	Human SH3P17 prote
20	1702	27.1	1097	22	ABW57814	Drosophila melanog
21	1536.5	24.5	936	22	ABG20576	Novel human diagno
22	1388.5	22.1	509	17	AAW05399	Human clone 65 pro
23	1361	21.7	990	22	AAU87256	Novel central nerv
24	1330.5	21.2	530	22	AAU17173	Novel signal trans
25	1183.5	18.9	436	22	AAW93738	Human protein sequ
26	1046	16.7	225	24	ABU71097	Human adipocyte Se
27	816.5	13.0	248	17	AAW05396	Human SH3P18 prote
28	774.5	12.4	182	22	AAO00069	Human polypeptide
29	712	11.4	164	23	ABP41930	Human ovarian anti
30	572.5	9.1	233	24	ABU71035	Human adipocyte Se
31	572	9.1	144	22	AAU87484	Novel central nerv
32	572	9.1	144	22	AAW43613	Human polypeptide
33	572	9.1	144	22	AAU19936	Novel human calcit
34	536.5	8.6	864	24	ABU03909	Human expressed pr
35	536.5	8.6	864	24	ABU03910	Human expressed pr
36	536.5	8.6	864	24	ABU03915	Human expressed pr
37	536.5	8.6	864	24	ABU03916	Human expressed pr
38	536.5	8.6	866	21	AAW86194	Nuclear transport
39	536.5	8.6	866	24	AAW03914	Human expressed pr
40	535.5	8.5	864	22	AAW40292	Human polypeptide
41	535.5	8.5	864	24	ABU03912	Human expressed pr
42	527	8.4	813	23	AAO20504	Protein of APP rel
43	503	8.0	754	22	ABW95697	Human protein sequ
44	503	8.0	754	24	ABU03911	Human expressed pr
45	486.5	7.8	896	17	AAW92750	Human EGF receptor

ALIGNMENTS

RESULT 1
ID AAV57444 strand: Protein; 1214 AA.
AAV57444;
AC AAV57444;
XX
XX
DT 28-FEB-2000 (first entry)
XX
XX
DE Mouse Eesl protein sequence.
XX
XX
KW Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
XX
OS Mus sp.
XX
XX
PN WO9955728-A2.
XX
XX
PD 04-NOV-1999.
XX
XX
PF 27-APR-1999; 99WO-CA00375.
XX
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
XX
PA (HSCR-) HSC RES & DEV LP.
XX
XX
PI Egan SE, Wang W, Sengar A;
XX
XX
DR WPI; 2000-052802/04.
DR N-PSDB; AA239008, AA239009.

PA (HSCR-) HSC RES & DEV LP.
 XX Egan SE, Wang W, Sengar A;
 XX MPI, 2000-052802/04.
 DR N-PSDB; AA239024, AA239025.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 XX infection
 PS Claim 14; Page 62-63; 99pp; English.
 XX
 CC The present invention specifically describes mammalian Ese1 and 2
 CC proteins (1) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (1) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Sep15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (1) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);
 CC sequences antisense to the (1) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (1) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (1) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Sep15 complex, then binding
 CC dynam to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese1 protein sequence.
 XX
 XX Sequence 1715 AA:
 SQ
 Query Match 99.8%; Score 6258.5; DB 21; Length 1715;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 481 LEALNDKQHQLQEGKIDIRCRLATOREIESTNKSRELRIATITLQOOLQOESQWGLRL 540
 DB 481 LEALNDKQHQLQEGKIDIRCRLATOREIESTNKSRELRIATITLQOOLQOESQWGLRL 540
 QY 541 IPEKQILSDQLKQVQNSLHSDSLTLTKRALEAKELARQQLREQLDEVERETRSKLOEID 600
 DB 541 IPEKQILSDQLKQVQNSLHSDSLTLTKRALEAKELARQQLREQLDEVERETRSKLOEID 600
 QY 601 VNNNQKLREIHSQQLQKRSLEAARLKQKEQERSKSELKQKEDARQVQENDKQWL 660
 DB 601 VNNNQKLREIHSQQLQKRSLEAARLKQKEQERSKSELKQKEDARQVQENDKQWL 660
 QY 661 EHVQEEOPRPKPKHEDELKREDSVRKKEAEERAKPEKQDSQSLFHHQBPATKATA 720
 DB 661 EHVQEEOPRPKPKHEDELKREDSVRKKEAEERAKPEKQDSQSLFHHQBPATKATA 720
 QY 721 PMSTTEKGPLTISAQESVYVYRALYPESHSHDEITIQPDIVWVDSQTEGPEMGLG 780
 DB 721 PMSTTEKGPLTISAQESVYVYRALYPESHSHDEITIQPDIVWVDSQTEGPEMGLG 780
 QY 781 ELKGTGMPANYAEKIPENEVPTPAKPYTDLTSAAPKLAIRETPAPLPTSSSEPTTP 840
 DB 781 ELKGTGMPANYAEKIPENEVPTPAKPYTDLTSAAPKLAIRETPAPLPTSSSEPTTP 840
 QY 841 NNMADPSSWPSSSNEKPEITDNDWDTMAQPSLTVPSAGQLRQSAFTPATGSSPSPV 900
 DB 841 NNMADPSSWPSSSNEKPEITDNDWDTMAQPSLTVPSAGQLRQSAFTPATGSSPSPV 900
 QY 901 GQGEKVEGLQALYPMRAKKNHNFNKSQVITVLEQDDMMWFGVQKGMFPKSYVK 960
 DB 901 GQGEKVEGLQALYPMRAKKNHNFNKSQVITVLEQDDMMWFGVQKGMFPKSYVK 960
 QY 961 LLSGPRKSTSIDTGTETSPASLKVYASPAKPAIPGEFFIMYTESSEQDGLTFQGD 1020
 DB 961 LLSGPRKSTSIDTGTETSPASLKVYASPAKPAIPGEFFIMYTESSEQDGLTFQGD 1020
 QY 1021 VIVVT-KKQDMMTGTVGDKSGVPPSNVYRLKDSGSGAGTKGSLGKKPEIAOVASYA 1079
 DB 1021 VIVVT-KKQDMMTGTVGDKSGVPPSNVYRLKDSGSGAGTKGSLGKKPEIAOVASYA 1079
 QY 1080 ATGPBQLTLAPQQLILIRKKNPGWMEGLQARGKKRQIGWPPANYVYKLLSPGSKITPT 1139
 DB 1080 ATGPBQLTLAPQQLILIRKKNPGWMEGLQARGKKRQIGWPPANYVYKLLSPGSKITPT 1139
 QY 1140 ELPKTAVPAYCOVIGMDYDTAQNDELAFSGQIINVLNKEDPDMWKGESGQGLFPS 1199
 DB 1140 ELPKTAVPAYCOVIGMDYDTAQNDELAFSGQIINVLNKEDPDMWKGESGQGLFPS 1199
 QY 1200 NYVKLTMDPSSQ 1213
 DB 1201 NYVKLTMDPSSQ 1214

RESULT 3
 AA32156
 ID AAY32156 standard; Protein; 1215 AA.
 XX
 AC AAY32156;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Human SH3D1A protein.
 XX
 KW SH3D1A; human; Down's syndrome; leukaemia; cancer;
 KW megakaryocytic abnormality; myeloproliferative disorder;
 KW platelet disorder; neural disorder; thrombocytopenia;
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
 XX
 OS Homo sapiens.
 OS
 PN WO953062-A2.
 XX

PD 21-OCT-1999.
 XX 16-APR-1999; 99WO-US08371.
 XX 16-APR-1998; 98US-0082007.
 XX (CEDA-) CEDARS SIGNAL HEALTH SYSTEM.
 PA Korenberg JR, Chen X;
 XX WP1, 1999-633829/54.
 DR N-PSDB; AAZ34572.
 XX
 PT Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 XX
 PS Claim 14; Fig 11; 99P; English.

CC This sequence represents the protein encoded by the human SH3D1A
 CC cDNA clone 11 (see AAZ34572). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukaemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AA32154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC chromosomal disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gain in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided: for suppressing
 CC cells unable to regulate themselves; screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality; myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.

CC Sequence 1215 AA;

Query Match 93.5%; Score 5860; DB 20; Length 1215;
 Best Local Similarity 92.9%; Pred. No. 0;
 Matches 1129; Conservative 39; Mismatches 45; Indels 2; Gaps 2;

```

QY 1 MAQPTTPGSGSLDVAVITVEERAKHDQQLSKPTAGFTTGDQANFFQSGLPQVLAQ 60
DB 1 MAQPTTPGSGSLDVAVITVEERAKHDQQLSKPTAGFTTGDQANFFQSGLPQVLAQ 60
QY 61 IWLADNMNDGMDQVEFSIAMKLIKLIKQYQLPSTLPVWKQOPVAISSAPARGIGI 120
DB 61 IWLADNMNDGMDQVEFSIAMKLIKLIKQYQLPSTLPVWKQOPVAISSAPARGIGI 120
QY 61 IWLADNMNDGMDQVEFSIAMKLIKLIKQYQLPSTLPVWKQOPVAISSAPARGIGI 120
DB 61 IWLADNMNDGMDQVEFSIAMKLIKLIKQYQLPSTLPVWKQOPVAISSAPARGIGI 120
QY 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLANGAPVITQPLPAFAHPAATWP 180
DB 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLANGAPVITQPLPAFAHPAATWP 180
QY 181 KSSSFRSGPGSGQLNTKQKQASPDVASAPAAEAAVQSSSLTKRQLFNSHDKTMSGHL 240
DB 181 KSSSFRSGPGSGQLNTKQKQASPDVASAPAAEAAVQSSSLTKRQLFNSHDKTMSGHL 240
QY 241 TGPQARTILMOSLTPQAOLASIMNLSIDIDQDKLTAEEFIILAMHLIDVAMSGQPLPVLP 300
DB 241 TGPQARTILMOSLTPQAOLASIMNLSIDIDQDKLTAEEFIILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPPSFRRAVSGSGMSVSSSVQDRLPEEPSSSEDEQOP-EKKLPVTFEDKKKENFER 359
DB 301 PEYIPPSFRRAVSGSGMSVSSSVQDRLPEEPSSSEDEQOP-EKKLPVTFEDKKKENFER 359
QY 360
DB 360

```

```

QY 360 GSVLEKRRQALTEQKREORLTAQLERAEOERKEREROEAKRQLELEKOLEKELE 419
DB 361 GNLEKRRQALTEQKREORLTAQLERAEOERKEREROEAKRQLELEKOLEKELE 420
QY 420 ROREERREKIEEREAARELERQOLEWERNRQELLNQRNKEQEGTVLAKAKTLEF 479
DB 421 ROREERREKIEEREAARELERQOLEWERNRQELLNQRNKEQEDIVLAKAKTLEF 480
QY 480 ELEALNDKQOLEBKODIRCLATQROETESTKSRRLTAETTHQQOOLEQSOQMLGR 539
DB 481 ELEALNDKQOLEBKODIRCLATQROETESTKSRRLTAETTHQQOOLEQSOQMLGR 540
QY 540 LIPERKQILSDQLKQVOONSLSHRDSLTLTKALAEKAPQALREQLDEVERETSXLQEI 599
DB 541 LIPERKQILSDQLKQVOONSLSHRDSLTLTKALAEKAPQALREQLDEVERETSXLQEI 600
QY 600 DVENNOLKELEIHSKQOLQKRSLEAARLKQKEERKSLELEKQEDAOQRVOERDKQW 659
DB 601 DIFNNQKLELEIHNKQOLQKRSMEAEARLKQKEERKILELEKQKEAQRRAQERDKQW 660
QY 660 LEHVQOE-EOPRPKPHEDRLKREDSVRKKEAEERAKPEMODQSLFHPHOBPAKLAT 718
DB 661 LEHVQOEDEHQRPKLHEBEKLRKEESVKKKQGEKKQOEADKGLRFLQHOBPAKPAV 720
QY 719 QAPSTTEKGPLTISAQESVYVYVYRALYPESRSHDEITIQPDIVWVDESQTEGPGL 778
DB 721 QAPSTAEKGPLTISAQENKVVYVYRALYPESHDEITIQPDIVWVDESQTEGPGL 780
QY 779 GGBELKGTGWFPANVYAEKIPENEVTPAKPVTDLTSAAPAPKLARETAPLPVTSSEPT 838
DB 781 GGBELKGTGWFPANVYAEKIPENEVTPAKPVTDLTSAAPAPKLARETAPLPVTSSEPT 840
QY 839 TPNWMAOFSSWPESSSNEKEPBTDMWDTMAAOPSLTIVSAQGLRQSAFTPATATGSSSP 898
DB 841 TPNWMAOFSSWPESSSNEKEPBTDMWDTMAAOPSLTIVSAQGLRQSAFTPATATGSSSP 900
QY 899 VLGGKEKEGLOAALTPWRAKDNHLPFNKSDVITYLEQDDMMWFEVQOKGMPKSY 958
DB 901 VLGGKEKEGLOAALTPWRAKDNHLPFNKSDVITYLEQDDMMWFEVQOKGMPKSY 960
QY 959 VKLISGPRKSTSIDTGPTESSPASLKEVASPAKPAIPGEEFIAMTYTESSEQDLTFQ 1018
DB 961 VKLISGPRKSTSIDTGPTESSPASLKEVASPAKPAIPGEEFIAMTYTESSEQDLTFQ 1020
QY 1019 GDVIVTTRKQGDMMTGTVGDKSGVFPNSYVRLKDSBESGTAQKSGCKREINQVASY 1078
DB 1021 GDVIVTTRKQGDMMTGTVGDAVFPNSYVRLKDSBESGTAQKSGCKREINQVASY 1080
QY 1079 AATGPEDLTAPGQLILIRKKNPGMWEGELQARCKKQIQGMFPANVYKLSPGTSKITP 1138
DB 1081 TATGPEDLTAPGQLILIRKKNPGMWEGELQARCKKQIQGMFPANVYKLSPGTSKITP 1140
QY 1139 TELPRTAVQPAVCQVIGMYDYTAQNDELAFSKQIINVLNKEDPDMMWKGVSQVGLFP 1198
DB 1141 TELPRTAVQPAVCQVIGMYDYTAQNDELAFENKQIINVLNKEDPDMMWKGVSQVGLFP 1200
QY 1199 SNVYKLTITDMDPSQ 1213
DB 1201 SNVYKLTITDMDPSQ 1215

```

RESULT 4

ID AA32155 standard; Protein, 1220 AA.

AA32155;

DT 01-FEB-2000 (first entry)

Human SH3D1A protein.

SH3D1A; human; Down's syndrome; leukaemia; cancer;
 megakaryocytic abnormality; myeloproliferative disorder;

KM platelet disorder; neural disorder; thrombocytopenia;
 KM haematopoietic disorder; cognitive dysfunction; microcephaly;
 KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy:
 OS Homo sapiens.
 XX MO9953062-A2.
 XX
 XX 21-OCT-1999.
 PD
 XX 16-APR-1999; 99MO-US08371.
 PF
 XX 16-APR-1998; 98US-0082007.
 PR
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
 PA
 XX Korenberg JR, Chen X;
 PI
 XX MPI; 1999-633829/54.
 DR
 XX N-PSDB; AA234571.
 XX
 XX Nucleic acid from the human SH3D1A gene and its products, useful for
 PT the diagnosis and treatment of myeloproliferative disorders and
 PT leukaemia
 PS
 XX Claim 14; Fig 9; 99pp; English.
 CC This sequence represents the protein encoded by the human SH3D1A
 CC cDNA clone 21 (see AA234571). SH3D1A contributes to the development
 CC of platelets and the pathogenesis of leukaemias, both in general
 CC and in particular those involving the megakaryocytic lineage. The
 CC gene maps to the small candidate region for low platelets on
 CC chromosome 21. The protein includes SH3 domains and EH domains,
 CC both associated with protein-protein interactions and the latter
 CC with maintenance of the cytoskeleton. At least 3 isoforms of
 CC SH3D1A exist (see AA232154-58). The invention provides methods for
 CC the diagnosis and treatment of megakaryocytic abnormality,
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,
 CC neural disorders, thrombocytopenia, platelet disorder on
 CC chromosome 21, low platelets in deletion for 21, association of
 CC gains in chromosome 21 with leukaemias, neural abnormalities,
 CC dysfunctions and disorders including brain malformations and
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,
 CC and colpocephaly. Methods are also provided for suppressing
 CC cells unable to regulate themselves, screening for a somatic
 CC alteration in the SH3D1A gene; monitoring the progress and
 CC adequacy of a treatment; monitoring tumour risk progress or
 CC megakaryocytic abnormality, myeloproliferative disorder,
 CC haematopoietic disorder, platelet disorder or leukaemia; and
 CC screening of drugs for cancer therapy.
 XX
 XX Sequence 1220 AA:
 SQ

Query Match 93.3%; Score 5847.5; DB 20; Length 1220;
 Best local Similarity 92.5%; Pred. No. 0;
 Matches 1159; Conservative 39; Mismatches 45; Indels 7; Gaps 3;

QY 1 MAQFPFPGSLDVAITVEERAKHQOFLSLKPIAGFITGDQARNFFPGSLPQPVLAQ 60
 DB 1 MAQFPFPGSLDVAITVEERAKHQOFLSLKPIAGFITGDQARNFFPGSLPQPVLAQ 60
 QY 61 IVALAMNNDGRDVOYEFSTAMKLIKXKOGYOLPSTLPVVMQOQVAISSAARFGIGI 120
 DB 61 IVALAMNNDGRDVOYEFSTAMKLIKXKOGYOLPSTLPVVMQOQVAISSAARFGIGI 120
 QY 121 ASNPPLTAAPVPMGSIPIVGMSPPLVSVPAAPVPLANGAPVPIQPLPAFAHPATWP 180
 DB 121 ASNPPLTAAPVPMGSIPIVGMSPPLVSVPAAPVPLANGAPVPIQPLPAFAHPATWP 180
 QY 181 KSSSFSRSGPGLNTKLOQAQSFVVASAPPAEAWAPPOSSRLKTRQLEFNSHDKTWSGHL 240
 DB 181 KSSSFSRSGPGLNTKLOQAQSFVVASAPPAEAWAPPOSSRLKTRQLEFNSHDKTWSGHL 240

QY 241 TGPARTIMOSSLPQAOASIAWNLSDIDQDGLTAEFFILAMHLIDVAMSGPLPVP 300
 DB 241 TGPARTIMOSSLPQAOASIAWNLSDIDQDGLTAEFFILAMHLIDVAMSGPLPVP 300
 QY 301 PEYIPPSFRVRSVSGMSVSISSSVDOURLPEEPPSSDEEQP-EKKLPVTFEDKRENFER 359
 DB 301 PEYIPPSFRVRSVSGMSVSISSSVDOURLPEEPPSSDEEQP-EKKLPVTFEDKRENFER 359
 QY 360 GSVELEKRRQALLQQRKEQERLAQLERAEQERKEBEROQERAKRQLELEKQERBE 419
 DB 361 GNLELEKRRQALLQQRKEQERLAQLERAEQERKEBEROQERAKRQLELEKQERBE 420
 QY 420 ROEEERREKREIRREAKKELEERORQLEWERRRROELNQRKKEQGVVLKRRRTLEF 479
 DB 421 RQREERREKREIRREAKKELEERORQLEWERRRROELNQRKKEQGVVLKRRRTLEF 480
 QY 480 ELEALNDKKHOLEGKLODIRCLATQROEIESTNKSRELRIAEITHLOQLOESQOMLGR 539
 DB 481 ELEALNDKKHOLEGKLODIRCLATQROEIESTNKSRELRIAEITHLOQLOESQOMLGR 540
 QY 540 LIPEKOILSDOLKQOVQNSLHRDLSLTLKRALEAKELAQOUREQLEVEREIRSKLOEI 599
 DB 541 LIPEKOILSDOLKQOVQNSLHRDLSLTLKRALEAKELAQOUREQLEVEREIRSKLOEI 600
 QY 600 DVFNNOQLKEIREHSKQOLQOKORSLAEARLKQKEORKSLELEKQEDARQROERDKQW 659
 DB 601 DVFNNOQLKEIREHSKQOLQOKORSLAEARLKQKEORKSLELEKQEDARQROERDKQW 660
 QY 660 LEHVQOE-EQPPRRKHEEDRLKREDSVYKKEAEERAKPEMODKQRLPHPHQEPKALAT 718
 DB 661 LEHVQOEEDHQPRKHEEKLRKREDSVYKKEAEERAKPEMODKQRLPHPHQEPKALAT 720
 QY 719 QAPWSTTEGRLTISAQESVKKVYTRALYPPFSRSDEITIQPGDIVM----VDESQFG 773
 DB 721 QAPWSTTEGRLTISAQESVKKVYTRALYPPFSRSDEITIQPGDIVM----VDESQFG 780
 QY 774 BEGMLGELGKGTGMPANVYAKIPENEVPTPAKPTDLSAPAPLARETPAPLPYVS 833
 DB 781 BEGMLGELGKGTGMPANVYAKIPENEVPTPAKPTDLSAPAPLARETPAPLPYVS 840
 QY 834 SEBSTTPNNMADPSSSTWSPSSNEKETDMDMAQPSLTVPSAGQLRORSAFTPATATG 893
 DB 841 SEBSTTPNNMADPSSSTWSPSSNEKETDMDMAQPSLTVPSAGQLRORSAFTPATATG 900
 QY 894 SSBSPVLGGEGVEGLQAOALYPMRAKKNHLPNFKSDVITYLEQDDMMFGEVQCKGM 953
 DB 901 SSBSPVLGGEGVEGLQAOALYPMRAKKNHLPNFKSDVITYLEQDDMMFGEVQCKGM 960
 QY 954 PKSVYKLSGPRKSTSIDGPTSPASLKRVASPAKPAIPIGEEFIAMTYESESQGD 1013
 DB 961 PKSVYKLSGPRKSTSIDGPTSPASLKRVASPAKPAIPIGEEFIAMTYESESQGD 1020
 QY 1014 LTFQOGDVIIVTKQGDMMTGVGDKSGVFPNNVYRLKDSGSGTAGTSGLKKEPIAQ 1073
 DB 1021 LTFQOGDVIIVTKQGDMMTGVGDKSGVFPNNVYRLKDSGSGTAGTSGLKKEPIAQ 1080
 QY 1074 VIASVATSPBEQTLTAPGQLILIRKKNPGCMMWEGELQARGKRRQIGWFPANVYKLISPGT 1133
 DB 1081 VIASVATSPBEQTLTAPGQLILIRKKNPGCMMWEGELQARGKRRQIGWFPANVYKLISPGT 1140
 QY 1134 SKITTELPKTAIVQAVCOVIMVYTYANDELAFSKQOILNVLNKEDPDMWKGVSQO 1193
 DB 1141 SKITTELPKTAIVQAVCOVIMVYTYANDELAFSKQOILNVLNKEDPDMWKGVSQO 1200
 QY 1194 VGLFSPNNYKLTITDMDPSQO 1213
 DB 1201 VGLFSPNNYKLTITDMDPSQO 1220

RESULT 5
 AA32154
 ID AA32154 standard; Protein; 1144 AA.

Sequence	1144 AA:	86.8%;	Score 5442.5;	DB 20;	Length 1144;
Query Match	86.8%;	Score 5442.5;	DB 20;	Length 1144;	
Best Local Similarity	87.2%;	Pred. No. 0;			
Matches 1060;	Conservative 37;	Mismatches 45;	Indels 73;	Gaps 3;	
QY	1	MAQFPPTPGSLDVAWLTVEERAKHDOFSLKPIAGITGDQARNFFPGSLPOPVLQAQ	60		
DB	1	MAQFPPTPGSLDVAWLTVEERAKHDOFSLKPIAGITGDQARNFFPGSLPOPVLQAQ	60		
QY	61	IMALDNMNDGRMDQVEFSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFIGCI	120		
DB	61	IMALDNMNDGRMDQVEFSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFIGCI	120		
QY	121	ASMPPLTVAAPVPMGSIPTVGMSPPLTVSVPAAPVPLANAPAVIQPLPAFAHPATWP	180		
DB	121	ASMPPLTVAAPVPMGSIPTVGMSPPLTVSVPAAPVPLANAPAVIQPLPAFAHPATWP	180		
QY	181	KSSFSRSGPSQSLNTKLQKQSPDVASAPPAEMAPQSSRLKYROLFNSHDTMSGHL	240		
DB	181	KSSFSRSGPSQSLNTKLQKQSPDVASAPPAEMAPQSSRLKYROLFNSHDTMSGHL	240		
QY	241	TGPOARTILMOSSLPOQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGOPLPVLP	300		
DB	241	TGPOARTILMOSSLPOQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGOPLPVLP	300		
QY	301	PEYIPSPSRRRRSSSGSVSSSSVDQRLPREPSESDQGP-EKKLPTEPDKKRENER	359		
DB	301	PEYIPSPSRRRRSSSGSVSSSSVDQRLPREPSESDQGP-EKKLPTEPDKKRENER	359		
QY	360	GSEVEKKRQALLTEQOREQERLQLEBAQERKERERQEOEAKROLELEKQEKOLE	419		
DB	361	GSEVEKKRQALLTEQOREQERLQLEBAQERKERERQEOEAKROLELEKQEKOLE	419		
QY	420	RQREERREKTEIRREBAKRELERQRLMEWENRRQELINQNKQEGTVLAKAKTLEF	479		
DB	421	RQREERREKTEIRREBAKRELERQRLMEWENRRQELINQNKQEGTVLAKAKTLEF	479		
QY	480	ELEALNDKQOLEKLODIRCLATQOQOESTKSKSELTIAETLTHLOOQLOESQOMGR	539		
DB	481	ELEALNDKQOLEKLODIRCLATQOQOESTKSKSELTIAETLTHLOOQLOESQOMGR	539		
QY	540	LIPKQILSDQKVOQNSLHRDLSLTLRKALEKELARQOLREQLDEVERETRSKLOEI	599		
DB	541	LIPKQILSDQKVOQNSLHRDLSLTLRKALEKELARQOLREQLDEVERETRSKLOEI	599		
QY	600	DVFNQNLKELREIHSKQOLKQNSLEAPARKQOEKQERKSLLEKQEKEDAQRRVQERDKQ	659		
DB	601	DVFNQNLKELREIHSKQOLKQNSLEAPARKQOEKQERKSLLEKQEKEDAQRRVQERDKQ	659		
QY	660	LEHVQOE-EQRPKRPHEEDRKXEDSVRKKEAEERAKPEWQDKOSRLFHHQOPAKLAT	718		
DB	661	LEHVQOE-EQRPKRPHEEDRKXEDSVRKKEAEERAKPEWQDKOSRLFHHQOPAKLAT	718		
QY	719	QAPWSTTEKGLTISAQESVKVYVYRLLYPFSRSHDEITIQPGDVIWVDSQTEGPEGM	778		
DB	721	QAPWSTTEKGLTISAQESVKVYVYRLLYPFSRSHDEITIQPGDVIWVDSQTEGPEGM	778		
QY	779	GSELKGTGMFPANYAEKIPENEPVTPAKPVTDLTSAAPAKLARETPAPLPVTSSSEST	838		
DB	781	GSELKGTGMFPANYAEKIPENEPVTPAKPVTDLTSAAPAKLARETPAPLPVTSSSEST	838		
QY	839	TPNNMADSSSPSSSNKEPTDMMDMAAPSLTIVSAGQLRQSAFPTATAGSSSP	898		
DB	841	TPNNMADSSSPSSSNKEPTDMMDMAAPSLTIVSAGQLRQSAFPTATAGSSSP	898		
QY	899	VLAGQEKTEGLOAQAALYFWRACKNHLNFNSADVIYVLQEQDMWFGEVQOQKMPKSY	958		
DB	901	VLAGQEKTEGLOAQAALYFWRACKNHLNFNSADVIYVLQEQDMWFGEVQOQKMPKSY	958		
QY	959	VKLISGPRKSTSIDTGPTEPSALAKVAPAAKRALPGEFFIAMYTESSEQDGLTQ	1018		
DB	959	VKLISGPRKSTSIDTGPTEPSALAKVAPAAKRALPGEFFIAMYTESSEQDGLTQ	1018		

QY 1019 GVIIVTKKDGWMTGTVDGKSGVFPSSNYRLKDSSESGTAGTSGSLGKKPEIAQVIASX 1078
Db 1001 -----EIAQVIASX 1009
QY 1079 AATGPEQLTLAPQGLILIRKKNFGWMEGELQARKKRQIGWFPANVYKLLSPGTSKIRP 1138
Db 1010 TATGPEQLTLAPQGLILIRKKNFGWMEGELQARKKRQIGWFPANVYKLLSPGTSKIRP 1069
QY 1139 TELPKTAVGPANQGVGMVDYTAQNDELAFSGQIINLNKEDPMWKGESGVGLRP 1198
Db 1070 TEPPKSTALAAVCQVIGMVDYTAQNDELAFNKGQIINLNKEDPMWKGESGVGLRP 1129
QY 1199 SNVYKLTMDPSQ 1213
Db 1130 SNVYKLTMDPSQ 1144

RESULT 6
AAU87168
ID AAU87168 standard; Protein; 1031 AA.
XX
AC AAU87168;
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #78.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227709.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-581633/65.

DR N-PSDB; ABK43498.

PT New isolated nucleic acid encoding a protein for diagnosing,
 PT preventing, treating or ameliorating medical conditions and used as
 PT food additives or preservatives -

PS Claim 9; SEQ ID No 686; 837bp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
 CC novel central nervous system protein. (I) and polypeptides (III) encoded
 CC by (I), are used to treat a medical conditions and in diagnosis of a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
 CC adenocarcinomas and irritable bowel syndrome, reproductive system
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
 CC acute kidney failure and blood related disorders e.g. myocardial
 CC infarction. The polypeptides can also be used to aid wound healing and
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
 CC maintain organs before transplantation, for supporting cell culture of
 CC primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 68.2%; Score 4277; DB 22; Length 1031;

Best Local Similarity 92.3%; Fred. No. 3.7e-266;

Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

Qy	1	MAQPTPPGSLDVMATVEERAKHQOQFSLKPIAGFTIGDQARFFFGSGLPQVLAQ	60
Db	8	MAQPTPPGSLDVMATVEERAKHQOQFSLKPIAGFTIGDQARFFFGSGLPQVLAQ	67
Qy	61	IMALADNMNDGRMDQVEFSIAMKLILKLGQYLPSTLPVWKQOPVAISSAPAFIGCI	120
Db	68	IMALADNMNDGRMDQVEFSIAMKLILKLGQYLPSTLPVWKQOPVAISSAPAFIGCI	127
Qy	121	ASMPPLTAAPVPMGSIPIVVGMSPPVSSVPRAAPVPLANGAPVQIPALPAFAHAPATWP	180
Db	128	ASMPPLTAAPVPMGSIPIVVGMSPPVSSVPRAAPVPLANGAPVQIPALPAFAHAPATWP	187
Qy	181	KSSSFSRSGSGSLNTKLQKQSPDVASAPPAEMAVPOSSRLKYNQLFNSHDKTMSGHL	240
Db	188	KSSSFSRSGSGSLNTKLQKQSPDVASAPPAEMAVPOSSRLKYNQLFNSHDKTMSGHL	247
Qy	241	T---GPOARTIMQSSLPQALASIMNLSIDIDQGLTAEEFLAMHLIDVAMSGQPLP	296
Db	248	TGSCIGPQARTIMQSSLPQALASIMNLSIDIDQGLTAEEFLAMHLIDVAMSGQPLP	307
Qy	297	PVLPPEYIPPSFRFVYSGMSVSISSSVQRLPEPSSSDEBOQ-EKLPVPEKKRE	355
Db	308	PVLPPEYIPPSFRFVYSGMSVSISSSVQRLPEEVEVEDEQOLEKLPVPEDEKRE	367
Qy	356	NFRSGVLEKRRQALLQOQRKEORLAQERAEORKEEREOEAKROLLELEKOLEQ	415
Db	368	NFRSGVLEKRRQALLQOQRKEORLAQERAEORKEEREOEAKROLLELEKOLEQ	427
Qy	416	RELEROEERERKEIERERAKLEEROROLEMERNRROGLNQRNKEGEGTVLXARRK	475
Db	428	RELEROEERERKEIERERAKLEEROROLEMERNRROGLNQRNKEGEGTVLXARRK	487
Qy	476	TLEFELALNDKKGHOLEGKQDTRCRLATROEIESTNKRELRIRAEITHLQOOLQESSQ	535
Db	488	TLEFELALNDKKGHOLEGKQDTRCRLATROEIESTNKRELRIRAEITHLQOOLQESSQ	547
Qy	536	MLGRILPEKQIISPOLKQVOONSILHRDLSLTLLKRALEAKELARQOLREQLDEVERETRSK	595
Db	548	MLGRILPEKQIISPOLKQVOONSILHRDLSLTLLKRALEAKELARQOLREQLDEVERETRSK	607
Qy	596	LOEIDVENNOLKELREIHSKQOLQOKRSLERARLKQKEQRKSLLEKQEKDAQRVQER	655
Db	608	LOEIDVENNOLKELREIHSKQOLQOKRSLERARLKQKEQRKSLLEKQEKDAQRVQER	667
Qy	656	DKQMLEHVQOE-EOPRRKHEEDRLKREDSYKKEAEERAKEMQOKSRLFFHPQEP	714
Db	668	DKQMLEHVQOE-EOPRRKHEEDRLKREDSYKKEAEERAKEMQOKSRLFFHPQEP	727
Qy	715	KLATQAWSTTEKGPLTISAOESYKVYYRALTYPFESRSHDEITIGDITVMDSEQTGE	774
Db	728	KLATQAWSTTEKGPLTISAOESYKVYYRALTYPFESRSHDEITIGDITVMDSEQTGE	787
Qy	775	PGWLGELKQKTGFWPANYAEKIPENEVPPPAKPVTDLTSAAPKALRETPAPLPTSS	834
Db	788	PGWLGELKQKTGFWPANYAEKIPENEVPPPAKPVTDLTSAAPKALRETPAPLPTSS	847
Qy	835	EPSTTPNNMADFSTTSSSSNEKETNNWTMAQPSLTYPSAQQLRQSAFTPATATGS	894
Db	848	EPSTTPNNMADFSTTSSSSNEKETNNWTMAQPSLTYPSAQQLRQSAFTPATATGS	907
Qy	895	SPSPVLQO 902	
Db	908	SPSPVLQO 915	

RESULT 7
AAM43519

ID AAM43519 standard; Protein; 1035 AA.
XX AAM43519;
XX
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 197.
XX
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; antiallergic; hepatotropic; antidiabetic;
KW antinflammatory; antitumor; anticonvulsant; antibacterial;
KW cardioparastic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human.
XX
OS Homo sapiens.
XX
XX WO200155308-A2.
XX
XX 02-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01309.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189674.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0246207.
PR 17-NOV-2000; 2000US-0246208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

```

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SW;
XX
XX WPI, 2001-488781/53.
XX N-PSDB; AAI63825.
XX
XX New isolated nucleic acids and polypeptides, useful for diagnosing,
XX treating and/or preventing human diseases and disorders -
XX
XX Claim 11; SEQ ID NO 197; 664pp + Sequence listing; English.
XX
XX The invention relates to human polynucleotides (AAI63803-AAI64012) and
XX the encoded proteins (AAM434497-AAM43660) useful for preventing, treating
XX or ameliorating medical conditions e.g. by protein or gene therapy. The
XX genes were isolated from a range of human tissues disclosed in the
XX specification. The nucleic acids, proteins, antibodies and (anti)agonists
XX are useful in the diagnosis, treatment and prevention of: (a) cancer,
XX e.g. breast and ovarian cancer and other cancers of the adrenal gland,
XX bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
XX urogenital; (b) immune disorders e.g. Addison's disease, allergies,
XX autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
XX Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1035 AA;
SQ
Query Match 68.2%; Score 4277; DB 22; Length 1035;
Best Local Similarity 92.3%; Pred No. 3.8e-266;
Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;
QY 1 MAQPTPTPGSLDVAIVTEERAKHDQQLSLKPIAGFTTGDQANFFQSGLPQPVLAQ 60
DB 8 MAQPTPTPGSLDVAIVTEERAKHDQQLSLKPIAGFTTGDQANFFQSGLPQPVLAQ 67
QY 61 IVALADNNNDGMDQVFSIANKLIKIKLQGYQLPSTLPVVKQOPVAISSAPAGIGCI 120
DB 68 IVALADNNNDGMDQVFSIANKLIKIKLQGYQLPSTLPVVKQOPVAISSAPAGIGCI 127
QY 121 ASMPPLTAVAPVPMGSIPIVGVSPPLVSVPPAAVPLANGAPVPIOTPLPAFHPAATMP 180
DB 128 ASMPPLTAVAPVPMGSIPIVGVSPPLVSVPPAAVPLANGAPVPIOTPLPAFHPAATMP 187
QY 181 KSSSFSGSGSGSLQTLQKQASFDVASAPAAEAVNPDSRLKRYROLFNSHDKTMSGHL 240
DB 188 KSSSFSGSGSGSLQTLQKQASFDVASAPAAEAVNPDSRLKRYROLFNSHDKTMSGHL 247
QY 241 T---GPOARTLMOGSLPQAOIASIWNLSIDIDQGLTAEEFIILAMHLIDVAMGQPLP 296
DB 248 TGSCGPGPQARTILMOGSLPQAOIASIWNLSIDIDQGLTAEEFIILAMHLIDVAMGQPLP 307
QY 297 PVLPEYIIPPSFRVRSSGSGSVISSSVQGLPEPSESEBQGP-EKKLPTFEDKKRE 355

```

```

DB 308 PVLPEYIIPPSFRVRSSGSGSVISSSVQGLPEPSESEBQGP-EKKLPTFEDKKRE 367
QY 356 NFERGSAVELEKRRQALLIQKQKEQERLAQLBRAQEKERBRQEQAKRQLELEKQLEKQ 415
DB 368 NFERGSAVELEKRRQALLIQKQKEQERLAQLBRAQEKERBRQEQAKRQLELEKQLEKQ 427
QY 416 RELEROREEERREKKEIRERBAKRELEROROLEMERBRNRQELNORNRKEQGVVLYKARK 475
DB 428 RELEROREEERREKKEIRERBAKRELEROROLEMERBRNRQELNORNRKEQGVVLYKARK 487
QY 476 TLFELBALNDKKHLEKGLQDIECRILATOROEIESTNKSRLEIRIAETIHLQOQLQESQ 535
DB 488 TLFELBALNDKKHLEKGLQDIECRILATOROEIESTNKSRLEIRIAETIHLQOQLQESQ 547
QY 536 MLGRLIPEKQILSLQKQVQONSILHRSLTLTKALPAKELARQQLAEQDVEVERTRSK 595
DB 548 MLGRLIPEKQILSLQKQVQONSILHRSLTLTKALPAKELARQQLAEQDVEVERTRSK 607
QY 596 LQETIDVNNOLKEIREIHSKQQLQKRSLEAARLKQKQERKSLELEKQEDAQRYQER 655
DB 608 LQETIDVNNOLKEIREIHSKQQLQKRSLEAARLKQKQERKSLELEKQEDAQRYQER 667
QY 656 DKQMLEHYQOE-EQPRPKPHEEDRLKREDSVRKKEAEERAKPEMODKOSRLFFPHQEP 714
DB 668 DKQMLEHYQOE-EQPRPKPHEEDRLKREDSVRKKEAEERAKPEMODKOSRLFFPHQEP 727
QY 715 KLAQAPWSTTEKQPLTISAOESVYVYRYALYFBSRSHDEITIQGDIYMWDESQGE 774
DB 728 KPAQVADWSTTEKQPLTISAOESVYVYRYALYFBSRSHDEITIQGDIYMWDESQGE 787
QY 775 PGWIGELKGTGTFPANYAEKIPENEVPPAPKAVTDLTSPAPKALRETPAPLPYTS 834
DB 788 PGWIGELKGTGTFPANYAEKIPENEVPPAPKAVTDLTSPAPKALRETPAPLPYTS 847
QY 835 EPSTTPNNMADFSSTPSSSNKEPBTNDWMAQPSLTVPSAQLRQSRSAFTPATATGS 894
DB 848 EPSTTPNNMADFSSTPSSSNKEPBTNDWMAQPSLTVPSAQLRQSRSAFTPATATGS 907
QY 895 SPSPVLQ 902
DB 908 SPSPVLQ 915

```

RESULT 8
ABG20575
ID ABG20575 standard; Protein; 932 AA.
AC ABG20575;
XX
XX 13-FEB-2002 (first entry)
XX
XX
XX Novel human diagnostic protein #20566.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
XX MO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI, 2001-639362/73.

DR N-PSDB; AAS84762.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 XX Claim 20; SEQ ID No 50934; 103pp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 932 AA:
 SQ
 Query Match 49.4%; Score 3097.5; DB 22; Length 932;
 Best Local Similarity 80.0%; Pred. No. 2.3e-190;
 Matches 603; Conservative 26; Mismatches 26; Indels 99; Gaps 3;
 QY 461 NNEQSTVVLKARKTLEFELEALNDKKQLQESKQODICRLATQOEISTTKSEELI 520
 DB 2 NNEQEDIVLVKAKKTLFELEALNDKKQLQESKQODICRLATQOEISTTKSEELI 61
 QY 521 AEITHLQOOLQESQOQLGRLIPEKQILSDQLKQVQNSLHRDLSILTKALAEKELARQO 580
 DB 62 AEITHLQOOLQESQOQLGRLIPEKQILSDQLKQVQNSLHRDLSILTKALAEKELARQO 121
 QY 581 LREQLDEVERETRSKIQEIDVFNNOIKELREIHSKQQLQKQSLERARLKQKEQEKSLIE 640
 DB 122 LRDQLDEVKETRSLQEIIDIFNNQIKELREIHNKQQLQKQSMERARLKQKEQEKSLIE 181
 QY 641 LEKQEDARVROERKQMLEHVQOE-EGPRPKPHEERLRKREDVRKKEAEERKPEM 699
 DB 182 LEKQEDARVROERKQMLEHVQOE-EGPRPKPHEERLRKREDVRKKEAEERKPEM 241
 QY 700 QDKQSLFPHQEPAPKATQAPWSTTEKPLTISAQESVAVVYRYALYFESRSHETI 759
 DB 242 QDKQSLFPHQEPAPKATQAPWSTTEKPLTISAQESVAVVYRYALYFESRSHETI 301
 QY 760 QGEDIWVDESQTEGEGMLGELIKGTGMPNANYAEKIPENEVPTPAKCVTDITSAAPK 819
 DB 302 QGEDIWVDESQTEGEGMLGELIKGTGMPNANYAEKIPENEVPTPAKCVTDITSAAPK 361
 QY 820 LALRETPAPLPYTSRPTTPNNMADPSSTWSSSNEKETPNWDMWAQPSLTVPSAQO 879
 DB 362 LALRETPAPLPYTSRPTTPNNMADPSSTWSSSNEKETPNWDMWAQPSLTVPSAQO 421
 QY 880 LKQSAFTPATATGSSPSPVLGGEGKEVGLQALTPMRAKKNHNFNKSQDITVLEQO 939
 DB 422 LKQSAFTPATATGSSPSPVLGGEGKEVGLQALTPMRAKKNHNFNKSQDITVLEQO 481
 QY 940 DMWPFEEVQOQKMFKPSYVKLISGVRKSTSIDTGPESPAKLKVASPAKPAIIGEE 999
 DB 482 DMWPFEEVQO-----E 491

QY 1000 FIAMTYSESQGDITFQGGDIVVTKKQDGMWTVGDKSGVPPSNVYRLKDSQSGSTA 1059
 DB 492 FIAMTYSESQGDITFQGGDIVVTKKQDGMWTVGDKSGVPPSNVYRLKDSQSGSTA 551
 QY 1060 GKTGSLGKKPEIAQVIAATGPEQTLTAPQQLILIRKQNGGWEGELQARGKKRQIG 1119
 DB 552 GKTGSLGKKPEIAQVIAATGPEQTLTAPQQLILIRKQNGGWEGELQARGKKRQIG 593
 QY 1120 WFPNANYKLLSPGTSKITPTELPTAVQPAVCQVIGMDYTDQNDELAFSGQIINVTN 1179
 DB 594 -----VCQVIGMDYTDQNDELAFNKGQIINVTN 623
 QY 1180 KEDPDMWKEVAGVGLPFPSTVYKLTMDPESQO 1213
 DB 624 KEDPDMWKEVAGVGLPFPSTVYKLTMDPESQO 657
 RESULT 9
 ID AAO17881 standard; Protein; 1681 AA.
 XX AAO17881;
 AC AAO17881;
 XX 22-AUG-2002 (first entry)
 DT
 XX Allergic disease examination method related human protein.
 DE Allergic disease; allergy; antiallergic; interestin 2; eosinophil;
 KW Allergic disease; allergy; antiallergic; interestin 2; eosinophil;
 KM atopic dermatitis; human.
 XX Homo sapiens.
 OS
 XX WO200233122-A1.
 PN
 XX 25-APR-2002.
 PD
 XX 11-OCT-2001; 2001WO-JP08937.
 PF
 XX 13-OCT-2000; 2000JP-0314093.
 PR
 XX (GENO-) GENOX RES INC.
 PA (NIGE-) JAPAN GEN NAT CHILDREN'S HOSPITAL.
 PA (EISA) EISAI CO LTD.
 XX Sugita Y, Hashida R, Ogawa K, Obayashi M, Nagaue T, Saito H;
 PI Takahashi E;
 DR N-PSDB; AAL47247.
 DR WPI; 2002-372313/40.
 XX Method for examining allergic diseases by differential display of
 PT intersectin 2 gene showing different expression particularly
 PT significant increase in eosinophils in patients
 XX
 XX Disclosure; Page 78-86; 90pp; Japanese.
 PS
 XX The present invention relates to a method for examining allergic diseases
 CC with intersectin 2 gene or a gene with equivalent function of intersectin
 CC 2 as an indicator gene, which comprises determining the expression level
 CC of the gene in the eosinophils in a patient, and comparing the expression
 CC level with that in the eosinophils of a healthy individual. The method is
 CC for examining allergic diseases, particularly atopic dermatitis, which is
 CC also applicable in screening candidate compounds for remedies. The
 CC present sequence is a human protein described in the exemplification of
 CC the invention.
 XX
 XX Sequence 1681 AA;
 SQ
 Query Match 49.3%; Score 3091.5; DB 23; Length 1681;
 Best Local Similarity 52.1%; Pred. No. 1.2e-189;
 Matches 656; Conservative 187; Mismatches 274; Indels 143; Gaps 29;
 QY 1 MAQPTPGGSLDWVAITVEERAKHDQFLSLKPTAGFITDQANFFQSGLPQPVLAQ 60

Db	13	MAQPFYANNGBNPNNWALTSEERTHDKRFJDLKSGSGYITTEDQARNFPLQGLPAVLAE	72
Qy	61	IMALADNMNDNRMDQVFSIAMIKLILKLQGYOLPSTLPVMKQOPV--AISSAPAFGIG	118
Db	73	IMALSDLNKDGKMQQSFISAMKLIKLQGOQOLPYULPRLMKQPMHSPILISA-RGWMG	131
Qy	119	GIASMPILITAVAPVPMG-----SIPVWGSPLIVSVPPPAVPLPANGAPV	165
Db	132	---SMPNLSISIQPLPPAPAPITLSGATSGTGLPLMPMPTPLVPVSTISLIP--NGTASL	185
Qy	166	IQPLPAFHAPAAWTWKSSESF--ASGPPSOULTLQOKXS--FDVASAP-----	211
Db	186	IQPLP-IDYSSSTLPHGSSYSISLMMGFg--GASIQKQOSLIDJGSSSSTSTASLSGNS	241
Qy	212	----AAEVAQSSRLKYROLFNSHDKTMSGHLTGPOARTILIMOSSLPLQALASIMWLS	267
Db	242	PKTGSBNAVAPQRLTKYRQKENTLIDKMSGVLISGFQARNMLQSNLSQTLATITWTLAD	301
Qy	268	IDQDGLTAEFIILAMHLIDVAMSGQPLPYLPPEYIIPPSFRARVSGSGMSVISSSDVO	327
Db	302	VDGDQOLAAEFIIILAMHLTDMAKAGQPLPLTLPPLVPSFR-----GKGQI--DSING	353
Qy	328	RLPEPSSSDQOPPKTLPYTPEDKKRNPERGSGTELEKRRQALLLEQKKEOELAQLER	387
Db	354	TLPSYQXNQ--BEEPQKTLPYTFEDKRXANYRGNNLEKRRQALMEQOQRAERAQOQEK	412
Qy	368	AAOEKEREBOEAMQROLEKOLEKORELEROEEERKRIERREAAKELEROLE	447
Db	413	EEWEKQOELQOEHKQOLEKLEKLEKRELEKREERKRDIERREAAKELERORLE	472
Qy	448	WERNROELLQNRNKEQEGTVVLKARRKTELELEBALNDKQOLEGKLQDIRCRLATQRO	507
Db	473	WERIROELLNQNRNEQEIYRLNSKKNLHLEBALNGKHQOISGRLODVLKMQOTKT	532
Qy	508	EIESNKSRELEIAITHLQOOLQSSQMLGRILPERKOIISDOLQOV--QNSLHRDSILT	566
Db	533	ELEVADKQCDIEIMIKOLOELOQOYOKLTYLVEKQDLNERIKMOPSTPDSVSL	592
Qy	567	LKRALAEKELARQOLREQDEVERETRSKLOEIDVFNNQLNELREIHSQOLOKORSLEA	626
Db	593	HKKSLEKELC-QRLKEQDIALEKTAKSLSMDSFNQOLNELRTTYTQOLAEQOLYKI	651
Qy	627	ARLKQOEORKSLELEKQ---EDAQRVERDRKQMLHVAQOEBOPRPRKHEDRLKR	682
Db	652	KRDKLKEIERKRLIMQKKLEDEAARAKQCKENLKENREKEBE-----EKQKROE	705
Qy	693	EDSVAKKAEEPAKPEMDKOSRLPHROEPAKLTQAPMSTTEKGLTISAQESVK--	739
Db	706	EKTQKIEEBEKKAEKQKQKD-----TLAAEKKRETA	740
Qy	740	--VVYVRLYPRESHSBEITIQPDIWVDESQTEPGMIGGELKSGTGMFPANYAEKI	797
Db	741	SVLVVYRALYPRFAARHDEMSFNSDITIQVDEKTYGEBFOMLYGSIQNGFMFPNYVERKM	800
Qy	798	P--ENEVPTPAKPVTDLTSAAPAKLALRETPALPVTSSBPTPNMNADEFSWTSSSN	855
Db	801	PESENE-----KAVSPKAL-----LPTVLSLATS-----TSSEPLSN	835
Qy	856	EKEPFDNMDTMAQPSLTVPSAGOLRQNSAFTPATATSSPSPVJGCEKYEGLOALY	915
Db	836	QPASVTDYON-VSPNLTVNTSMQ--KKSAPFRTVSPG-SVSPHGOQJVENILKAQALC	891
Qy	916	PMRAKDDHNLNKSQDIVTLEQODMMNGEYQOGKGMFPKSVYVLISGPVRKXSIDTG	975
Db	892	STAKKDDHNLNKSQDITVLEQOQNMFMGEVHGGRGMPKSVYKIPGSEYK-----	944
Qy	976	PTESPASILKRVAS--PAAKPAIPGEEFIAMTYESESQGDILTFQOGADVIYVTKQGDWMT	1033
Db	945	-REEBEALYAAVNNKPFSAAYSGVEEYIALYIYSSVEGODLTFTEGEEIYVTKQGDWMT	1003
Qy	1034	GTVDKSGVFPSPSNVYRLDSESSGAGTKGSLGKKPEIQAQVILASVYATGPEQTLAPQOL	1093

Db	1004	GSIDRSGISFFSNVYVKRDQESFSSASAKSGASGNKKPEIAQVTSAYVSGSEQLSLAARGQL	1063
Qy	1094	ILIRKKNPGCWEGEGLQARGKKRQILGFPANVYKLSPGTSIKTPTTELPTAVQAVCOV	1153
Db	1064	ILILKKNKTSIGMWQGLQARGKKRQKGMFPASHVYKLLGPSSESRATPAHP-----VCQV	1116
Qy	1154	IGMVDYTAQNDELAFSGKQILINLNKEDDPMMWCGEVSQGVLEPPSNVYKLLTDMDSQO	1213
Db	1117	IAMVDYANNDELSFSGKQILINWNRKDDPDMMWQGEINVTGLFPPSNVYKMTTSDPSQO	1176
RESULT 10			
Id	AAV32157	standard; Protein; 648 AA.	
XX	AAV32157;		
AC	AAV32157;		
XX			
DT	01-FEB-2000	(first entry)	
XX			
DE	Human SH3D1A protein.		
XX			
KW	SH3D1A; human; Down's syndrome; leukaemia; cancer;		
KW	megakaryocytic abnormality; myeloproliferative disorder;		
KW	platelet disorder; neural disorder; thrombocytopenia;		
KW	haematopoietic disorder; cognitive dysfunction; microcephaly;		
KW	lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.		
XX			
OS	Homo sapiens.		
XX			
PN	W09553062-A2.		
XX			
PD	21-OCT-1999.		
XX			
PF	16-APR-1999; 99WO-US08371.		
XX			
PR	16-APR-1998; 98US-0082007.		
XX			
PA	(CEDA-) CEDARS SINAI HEALTH SYSTEM.		
XX			
PI	Korenberg JR, Chen X;		
XX			
DR	WPI; 1999-633829/54.		
DR	N-PDB; AA234573.		
PT	Nucleic acid from the human SH3D1A gene and its products, useful for		
PT	the diagnosis and treatment of myeloproliferative disorders and		
PT	leukaemia		
PS	Claim 14; Fig 13; 99pp; English.		
XX			
XX	This sequence represents the protein encoded by the human SH3D1A		
CC	CDNA clone 5 (see AA234573). SH3D1A contributes to the development		
CC	of platelets and the pathogenesis of leukaemias, both in general		
CC	and in particular those involving the megakaryocytic lineage. The		
CC	gene maps to the small candidate region for low platelets on		
CC	chromosome 21. The protein includes SH3 domains and EH domains,		
CC	both associated with protein-protein interactions and the latter		
CC	with maintenance of the cytoskeleton. At least 3 isoforms of		
CC	SH3D1A exist (see AAU32154-58). The invention provides methods for		
CC	the diagnosis and treatment of megakaryocytic abnormality,		
CC	myeloproliferative disorder, platelet disorder, acute leukaemia,		
CC	neural disorders, thrombocytopenia, platelet disorder on		
CC	chromosome 21, low platelets in deletion for 21, association of		
CC	gains in chromosome 21 with leukaemias, neural abnormalities,		
CC	dysfunctions and disorders including brain malformations and		
CC	corresponding cognitive dysfunctions, microcephaly, lissencephaly,		
CC	and colpocephaly. Methods are also provided: for suppressing		
CC	cells unable to regulate themselves; screening for a somatic		
CC	alteration in the SH3D1A gene; monitoring the progress and		
CC	adequacy of a treatment; monitoring tumour risk progress or		
CC	megakaryocytic abnormality; myeloproliferative disorder,		
CC	haematopoietic disorder, platelet disorder or leukaemia; and		
CC	screening of drugs for cancer therapy.		

XX Sequence 648 AA;
SQ Query Match 49.0%; Score 3070.5; DB 20; Length 648;
Best Local Similarity 94.1%; Pred. No. 7,4e-189;
Matches 609; Conservative 20; Mismatches 17; Indels 1; Gaps 1;
QY 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFFGSGLPQVLAQ 60
DB 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFFGSGLPQVLAQ 60
QY 61 IVALADNMNDGRMDQVEFSIAMKLIKLIQGYLPSTLPPVMKQPVASISAPAFIGIG 120
DB 61 IVALADNMNDGRMDQVEFSIAMKLIKLIQGYLPSTLPPVMKQPVASISAPAFIGIG 120
QY 121 ASMPPLTAAVAVPMGSIPIVVGMSPTLVSSVPTAAVPEPLANGAPVPIQPLPAFAHPAATLP 180
DB 121 ASMPPLTAAVAVPMGSIPIVVGMSPTLVSSVPTAAVPEPLANGAPVPIQPLPAFAHPAATLP 180
QY 181 KSSSFSGSGSGSLQNTKLOKAGSFDVASAPPAEMVAVPOSSRLKTRQLFNSHDKTMSGH 240
DB 181 KSSSFSGSGSGSLQNTKLOKAGSFDVASAPPAEMVAVPOSSRLKTRQLFNSHDKTMSGH 240
QY 241 TGFQARTILMOSSLPOAQLASINWLSDDIDQKLTAEFPLAMHLIDVAMSGPLPPVLP 300
DB 241 TGFQARTILMOSSLPOAQLASINWLSDDIDQKLTAEFPLAMHLIDVAMSGPLPPVLP 300
QY 301 PEPIPSFRVRVSGSGMSVYSSSVQRLPEEPPSSSEDBQOP-EKULPVTFEDKKRNFEP 359
DB 301 PEPIPSFRVRVSGSGMSVYSSSVQRLPEEPPSSSEDBQOP-EKULPVTFEDKKRNFEP 359
QY 360 GSVELEKRRQALLQQRKEGERLAOLERAQERKERERQERQERQERQERQERQERQER 419
DB 360 GSVELEKRRQALLQQRKEGERLAOLERAQERKERERQERQERQERQERQERQERQER 419
QY 420 ROREERRRKEIERREAKELERQOLEWERNRROELNORKEOEGTVLTKARKTLEF 479
DB 420 ROREERRRKEIERREAKELERQOLEWERNRROELNORKEOEGTVLTKARKTLEF 479
QY 480 ELBALNDKKGQLEGKQDTRCRLATOROBIESTNKREIRIAITHTLQOOLQESQOMLGR 539
DB 480 ELBALNDKKGQLEGKQDTRCRLATOROBIESTNKREIRIAITHTLQOOLQESQOMLGR 539
QY 540 LIPKQILSDQKQVOONSILHRPSLTLTKRALEAKELAQOLDEQDEVERETRSLOEF 599
DB 540 LIPKQILSDQKQVOONSILHRPSLTLTKRALEAKELAQOLDEQDEVERETRSLOEF 599
QY 600 DVFNNOQLKEIREIHSKQQLQKRSLEAARLKQERKSLLEKQKE 646
DB 600 DVFNNOQLKEIREIHSKQQLQKRSLEAARLKQERKSLLEKQKE 646
QY 601 DIFNQQLKEIREIHSKQQLQKRSLEAARLKQERKSLLEKQKE 647
DB 601 DIFNQQLKEIREIHSKQQLQKRSLEAARLKQERKSLLEKQKE 647
RESULT 11
AAM79199
ID AAM79199 standard; Protein; 1697 AA.
XX AAM79199;
XX 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 1861.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX
XX MO200157190-A2.
XX 09-AUG-2001.
XX

PF 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang Y, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52332.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
XX
XX useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 4246-4249; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM60020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX SQ Sequence 1697 AA;
XX
XX Query Match 48.9%; Score 3068; DB 22; Length 1697;
XX Best Local Similarity 51.0%; Pred. No. 4,1e-188;
XX Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;
QY 1 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFFGSGLPQVLAQ 60
DB 2 MAQFPFGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARFFFGSGLPQVLAQ 61
QY 61 IVALADNMNDGRMDQVEFSIAMKLIKLIQGYLPSTLPPVMKQPV-AISSAPAFIGIG 118
DB 62 IVALADNMNDGRMDQVEFSIAMKLIKLIQGYLPSTLPPVMKQPV-AISSAPAFIGIG 120
QY 119 GIASMPPLTAAVAVPMGSIPIVVGMSPTLVSSVPTAAVPEPLANGAPVPIQPLPAFAHPAATLP 165
DB 121 ---SMENSLIPPLPAADITSLSSATSGTNPPLMPTPLVPSTSSLP---NGTASL 174
QY 166 IOPPLPAFAHPAATWPKSSSFS--RSGPSGQLTKLOKAGS-PDVASAP----- 211
DB 175 IOPPL-IPYSSSTLPHGSSYSILMWGFG---GASIQKQSLDLSSSSSTSTASLGSNS 230
QY 212 ----AAEMAVPOSSSLKTRQLFNSHDKTMSGHITGPQARTILMOSSLPOAQLASINWLSDD 267
DB 231 PRTGISEMAVPOPTRLKTRQKQNTLDKMSGYLSFGQARNALQNTLQQTATATWTLAD 290
QY 268 IQDGKLTAEFPLAMHLIDVAMSGOPLPPVLPPEPIPSFRVRVSGSGMSVYSSSVQD 327
DB 291 VPDGQQLAEFPLAMHLIDVAMSGOPLPPVLPPEPIPSFRVRVSGSGMSVYSSSVQD 342
QY 328 RLPEPPSSSEDBQOP-EKULPVTFEDKKRNFEPGSGVELEKRRQALLQQRKEGERLAOLER 387
DB 343 TLPSYQXQG-EPEPKULPVTFEDKKRNFEPGSGVELEKRRQALLQQRKEGERLAOLER 401

```

QY 388 AEOEKEREROEAKROLELEKOLEKOREREROREERREARERREAREROROLE 447
DB 402 EEWEEKOROLEKOLEKOLEKOREREROREERREARERREAREROROLE 461
QY 448 WERNROELNORNEOGTVLAKARKTLEFELEBALNDKHKOLEGKLODRICRATORO 507
DB 462 MERIRGOLLNORNEOEIVRLNKKKNIHLEBALNGKQISGRQDVRILKMQOTKT 521
QY 508 EIESTNKRELEIAITLHQOOLQESQOMGLRIEIKOILSDQKQV-ONSLSHDSILT 566
DB 522 ELEVLDKOCODLEIMEIKOLOELQEQNKLYLVEBEKOLMERIKMFSNTPDGVSILT 581
QY 567 LKBALEAELARQOLREQLDEVERRETSKLOEIDVFNNO----- 605
DB 582 HKKSLEKEBELC-QRLKEQDLDALEKETASKLSBMSFNQOLKCGNMDSVLOCILSLSL 640
QY 606 -----LKELEREHSKOOLQOKORSLAARLKOKEORERKLEKOK-----EDAPORVOER 655
DB 641 NNLFLLLELRETVNTQOLALEQYKIRKDKLKEIERKRLBELMOKKLEDEARAKQGX 700
QY 656 DKOMLEHVQOEOPPRKPHEDRLKREDSVRKKAERAKPEMODKOSRLFHPQEPAX 715
DB 701 ENLMKERNRKEBE-----EKOKRLOEKETOKEIOEBERKKEKOKRDXD----- 744
QY 716 LATQAPWSTTEKGPLTISAOSSVK-----VYYRALYEPESRSHDEITOPGDIVMDES 770
DB 745 -----TLKAEKKRETASVLYNTRALYFEANNDHMSFNSDIILQVDEK 789
QY 771 QTEBERGMLGELKGTGWFNPANYAEKIP--ENEVPTPAKPYDILTASAPKILARETPAP 828
DB 790 TYGEQGMVLYGSGFGWFGWPCVYVEKMSSENE-----KAVSPKKAL----- 831
QY 829 LPTVTSSESTTPNMADPSSSTWPSSENEKPEETDMMDTAAOBSLTVPBAGOLRQASATP 888
DB 832 LPTVTSLSATS-----TSSEPLSSNQPASVTDYON--VFSNLTVTNVTSMQ--KKSAPTR 881
QY 889 ATATSSSPVLYGQEKVEGLOAQLALYMPRAKONHLNPNKSDVTTVLEQODMFGFVQ 948
DB 882 TVSPG-SVSPHGGQOVENLKAQMLCSWTAKKDHNLNFSKRDITVLEQDENMFGFVH 940
QY 949 GQKGMFPKSYVYLISGPVRKSTSIDTGTSPASILKRVAS--PAKPAIPGEEPTAMTY 1006
DB 941 GGRGMFPKSYVKIIFGSEVK-----REBEALYAIVNKKPISAASVEEYIALPY 992
QY 1007 ESSEGGDILTPQOGDIVVTTKODGMWGTGVDKSGVPPSNRYRLKDSGSGTAGTSGLG 1066
DB 993 SSVEPGDLTFTEGEEILVTOKDGEMWMTSGIDRSGIFPSNRYKPPDOESFGSASKSGASN 1052
QY 1067 KKPEIAOVIASAAATGPEQLTLAPGOLLIRKONGGMEGLQARGKKQIGMFPANVY 1126
DB 1053 KKPEIAOVTSAVAVASGSEQLSLAPGOLLILKNGSGMWGSLQARGKKRQGMFPASHV 1112
QY 1127 KLSSEGTSKITPTLEPKTAVOPAVQOVIGMYDYTAQNDELAFSGQIINVLNKEDPDWM 1186
DB 1113 KILGSSSRATPAPHP-----VCQVIAWMDYANNDDELSEFSKQILINWKNODPDWM 1165
QY 1187 KGEVSGQVGLPPSNRYKLTMDPSQO 1213
DB 1166 QGEINGVTGLPPSNRYKMTTDSPSQO 1192

```

```

KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
OS Mus sp.
XX MO9955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99WO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX 05-FEB-1999; 99US-0118739.
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Wang W, Sengar A;
XX WPI; 2000-052802/04.
XX N-PSDB; AA239010, AA239011.
XX
XX New nucleic acid encoding Ees1 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection.
XX
XX Claim 33; Page 48; 99pp; English.
XX
XX The present sequence represents mouse Ees2. The present invention
XX specifically describes mammalian Ees1 and 2 proteins (I) and their splice
XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
XX are involved in regulation of clathrin-mediated endocytosis (as a complex
XX with Ees1s protein), vesicular trafficking and actin cytoskeleton.
XX Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
XX mutants); (II)-specific antibodies (Ab); sequences antisense to the (I)
XX polynucleotide; agents that downregulate expression of Ees genes or
XX antagonists of an Ees binding partner are used to treat diseases
XX associated with undesirable endocytosis and resulting changes in cellular
XX function. Particularly overexpression of Ees1 is used to block
XX clathrin-mediated endocytosis in vivo or in cell cultures, while
XX administration of (I) is used to promote endocytosis of selected cells.
XX (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
XX cells that can be stimulated to proliferate by a growth factor receptor;
XX and similar compounds (also inactive Ees mutants) can be used to prevent
XX viral infection. Endocytosis may also be regulated, in vivo or in cell
XX cultures, by forming an Ees-Ees1s complex, then binding dynam to the
XX complex. Generally conditions that can be treated include cancer;
XX abnormal cell division or migration; viral infection; or abnormal
XX receptor signalling, tissue development or synaptic transmission.
XX
XX Sequence 1197 AA;
SQ
Query Match 48.6%; Score 3049.5; DB 21; Length 1197;
Best Local Similarity 52.4%; Pred. No. 3.9e-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
QY 1 MAORPTFGSLDVMTATVEERAKHDQOFLSLKIAEPFINGDQARNFFQSGLPVLAQ 60
DB 1 MAQPTAMNGGPNWMTATSEBTRHDQFNLKSGGYITGDQRTFLDGLAPVLA 60
QY 61 IVALADNNDGRMDQVEFSIAMKLKIKLOGYQLPSTLPVPMKQOPV--AISSAPAGIG 118
DB 61 IVALSDLNKQKMDQVEFSIAMKLKIKLOGYQLPSTLPVPMKQOPV--AISSAPAGIG 119
QY 119 GIASMP-----PLTAVAPV--PMG-----SIPVVGMSPPVLSVPPPAVPLIANGAP 164
DB 120 ---SMPNLSIHQPLPVPAPATPLSATSGTSIPPLMMPAPLVSVSTSLP---NGTAS 173
QY 165 VIQPLPAFAHRAATWPSSSFS--RSGPQGLNKKQKQAS--PDVAAPV----- 211
DB 174 LIQPL-SIPYSSSTLPHASSSTLMGGFG--GASIQKQSLIDLQSSSSSTSTASLSGN 229
QY 212 -----AAEWAVPQSSRLKQLFNSHDKTMSGHLTGQARTILMQSSLPQQLASIWNL 266

```

RESULT 12
AA157445
ID AA157445 standard; Protein; 1197 AA.
XX
XX AA157445;
XX
XX 28-FEB-2000 (first entry)
XX
XX Mouse Ees2 protein sequence.
XX
XX Mouse; murine; Ees1; Ees2; endocytosis; vesicular trafficking;

```

Db      230  SPKGTSEWAVPQPSRLKTRQKENSILDKGWSYLSGFOANALQSNLSOTQATITWTLA 289
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      267  DIDDOCKLAEETILMHLIDVAMSGOPLPVLPPVIPPSPFRVSSGSMYSISSSD 326
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      290  DIDGDDQLAEETILMHLIDVAMSGOPLPVLPPVIPPSPFRVSSGSMYSISSSD 341
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      327  QRLPEPSSSEDOQPEKLPVTFEDKKRENFERSVLEKRRQALLQOORKEOERLAQ 386
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      342  GILPSPQKQ-EEEPQKLPPTTEDGRKANFERGNELEKRRQVLMQOOREERAKQ 400
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      387  RAEOERKERERQOAKROLEKOLEKORERERERERERERERERERERERERER 446
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      401  KEEMERKORELOQEWKQOLEKLEKLEKORERERERERERERERERERERER 460
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      447  EWERNRROELINORNEOEGTVLKKARKTLEFELALNDKQOLEGKODIFCRILAT 506
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      461  EWERLRROELISQTEEDIVRLSSRKSLHLEBAVNGKHQOISGRLODVQIRKQ 520
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      507  OEIESTNKSRELIATITLQOOLQOOLQOOLQOOLQOOLQOOLQOOLQOOL 566
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      521  TELFVLDKODLEIMIKOQLEKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 580
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      567  LKRALAEKELARQOLREOLDEVERETRSKLOEIDVENNOQKELREIHSKQOL 626
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      581  LHKSESEKEELCOQLKEQDLEKETAASKLSEMSFNNOQKELREIHSKQOL 640
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      627  ARLKQOEORKSLE-LEKQ--EDAKRQOEKQKQKQKQKQKQKQKQKQKQK 682
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      641  KQDKLEIERKRLQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 694
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      683  EDSVRKEAEERAKPEMODKOSRLPHQEPALATQAPMSTTEKQPLTISAQESK 742
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      695  EKSQDKTOEER--KAEAKQSE-----TASA-----LVN 721
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      743  YRALYFESRSHDEITOPEDIYVNDSEOTGEPMGLGELGKGTGMPFANYAEKI 802
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      722  YRALYFPEARNDHEMFSSGDIQVDEKTVGEGWLYGSGQKGFPCVYEVKLSSE 780
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      803  PTPAKVNTDLSAPAKALRETPAPLPVTSSEPTTPNNMADPSSSTWSSSEKETE 862
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      781  -----KALSFKKALLPPTVSLSATSTSSQPPASVTDYHNV--SFSNLTVNT 824
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      863  WPTMAQPSLTVPSAQQLRQSAFTPATATGSSPSVVGGEVEGLQOALYPMRAX 922
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      825  --TW-----OQKSAFTTVPFG-SVSPTHGQQAVENTLKQALCSWPAKXB 867
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      923  NHIENFKSDVITVLEQODMMWFEVQOGKGMFPKSYVYLISG-PVRS--TSID 979
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      868  NHIENFKSDVITVLEQODMMWFEVQOGKGMFPKSYVYLISG-PVRS--TSID 927
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      980  PALSKEVASPAKPAIP-GEETIAMTYESSGEGDILTPQOGDVIIVTKKGDWMT 1038
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      928  PVS---TAAVPISTAPVGEDYIALYSYSVEBPDILTFEGEILVLTQKGEWMT 984
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      1039  KSGVPSNVYRLKDSGSGTAGTGLGKKEPETAUITYASAAAGPQOLTLAPQOLI 1098
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      985  RTGIFPSNVYRPQDQENFGASGASNNKPEITAQVTAASAGTQOLSLAPQOLI 1044
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      1099  KNPFGWEGELQARKKROIGMPFANYVYKLSPTGKITPTTELPKTAPOVAVQ 1158
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1045  KNTSGWQEGELQARKKROIGMPFANYVYKLSPTGKITPTTELPKTAPOVAV 1097
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      1159  YTAQNDDELAFSGQIITVANKEDPMMWKEVSGQVGLPSPNYVYKLTMDPSQ 1213
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1098  YMANNEDELAFSGQIITVANKEDPMMWKEVSGQVGLPSPNYVYKLTMDPSQ 1152
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT 13
AA57450
ID AA57450 standard; Protein; 1658 AA.
XX
AC AA57450;

```

XX      28-FEB-2000 (first entry)
DT      Mouse Ees2L protein sequence.
XX      Mouse; murine; Ees2; endocytosis; vesicular trafficking;
XX      regulation; actin cytoskeleton; detection; cancer; infection;
XX      EH-domain and SH3-domain regulator of endocytosis; anticancer;
XX      antiproliferative; antiviral.
OS      Mus sp.
XX      MO995728-A2.
XX      04-NOV-1999.
XX      27-APR-1999; 99MO-CA00375.
XX      27-APR-1999; 98CA-223021.
XX      05-FEB-1999; 99US-0116739.
XX      (HSCR-) HSC RES & DEV LP.
XX      Egan SE, Wang W, Sengar A;
XX      MPI; 2000-052802/04.
XX      DR      N-PSDB; AA239026, AA239027.
XX      New nucleic acid encoding Ees1 and 2 proteins, involved in regulation
XX      of endocytosis, used e.g. for treating cancer or preventing viral
XX      infection.
XX      Claim 33; Page 69-70; 99p; English.
XX      The present invention specifically describes mammalian Ees1 and 2
XX      proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
XX      regulator of endocytosis). (I) are involved in regulation of clathrin-
XX      mediated endocytosis (as a complex with Ees1 protein), vesicular
XX      trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
XX      mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
XX      sequences antisense to the (I) polynucleotide; agents that downregulate
XX      expression of Ees genes or antagonists of an Ees binding partner are
XX      used to treat diseases associated with undesirable endocytosis and
XX      resulting changes in cellular function. Particularly overexpression of
XX      Ees1 is used to block clathrin-mediated endocytosis in vivo or in cell
XX      cultures, while administration of (I) or Ab are used to suppress abnormal
XX      proliferation of cells that can be stimulated to proliferate by a growth
XX      factor receptor; and similar compounds (also inactive Ees mutants) can be
XX      used to prevent viral infection. Endocytosis may also be regulated, in
XX      vivo or in cell cultures, by forming an Ese-Ees1 complex, then binding
XX      dynamitin to the complex. Generally conditions that can be treated include
XX      cancer; abnormal cell division or migration; viral infection; or abnormal
XX      receptor signalling, tissue development or synaptic transmission. The
XX      present sequence represents mouse Ees2L protein sequence.
XX      Sequence 1658 AA;
XX      Query Match 48.6%; Score 3049.5; DB 21; Length 1658;
XX      Best local similarity 52.4%; Pred. No. 6.1e-187;
XX      Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
Qy      1 MAQFPFPGSGSLDWAITYEERAKIDQFLSKPIAGFTTGQARNFFQSGLPQVLAQ 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MAQFPFPGSGSLDWAITYEERAKIDQFLSKPIAGFTTGQARNFFQSGLPQVLA 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      61 IVALADMMNDGMDQVSESIAMKLIKIKLQGLQGLPVTLPPIKQPMPSPLISA-RFG 118
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 IVALADMMNDGMDQVSESIAMKLIKIKLQGLQGLPVTLPPIKQPMPSPLISA-RFG 119
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      119 GIASMP-----PLTAVAVP--PMG-----SIPVGSMPPLVSSVPPAAPPPLANG 164
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      120 ---SMPNLSTHQPLPVPVAVIATPPLSSATSGTSTIPPLMMPAFLVPSVSTSLP---NGTAS 173
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

165 VIOPLPAFAHPAAWPKSSSFS--RSGPSQLNTKLOKAO--FDVYASAP----- 211
174 LIQPL-STIPSSSTIPHASSTYSLMMGFG--GASIQQAQSLIDGSSSTSSSTASLSCN 229
212 -----AAEMAVPOSSRLKYROLFNSHSDTMSGHLTGPARITLMQSSLPQOLAISTWNLS 266
230 SPKGTSSMAVQPQRLKYRQKFNSLDKMGSGYLSGFAANALQSNLSQQLATITWLA 289
267 DIDQDKLTAEFFILAMHLIDVANSQPLPPLPPEYTPPSRRRRSSSGMSVSSSSVD 326
290 DIDGGQKABEFFILAMHLIDMAKAGQPLPLPPELPPEPSRR-----GGQV--DSVN 341
327 QRLPEPSSEDOQEKLPVTFEDKKENFERGVELEKROALLQEOQKOEATLALE 386
342 GTLPYQKTO--EEFQKLPPTFEKRYANERGMELERKQVLMEOQOEAEKKAQKE 400
367 RAEQERKERQOEQAKQOLEKQLEKQRELERQREERKEIERREBAKRELERQOL 446
401 KEEMERKQRELQOEQWKQLEKQLEKQRELERQREERKEIERREBAKQELERQRL 460
447 EEMENRRQELLNQRKQOEQVYVYKARKTLEFELEALNDKKHOLEGLQDRCRLATQ 506
461 EEMERLRQELLQKTRQOEQDIYVLSRRKSLHLEAVNGHQQISGRLQDVQIRKQTK 520
507 QEISTNSRELRIETHTLQOOLQESQOMGLPPEKQIISDOLKQVQNSLHRDLSLT 566
521 TELEVLQKQDEIETIKQOLEKQYQNKLYLPEKQLEMERIKNNQSLNTPDSGSL 580
567 LKRALAEKELARQOLQDEVERETRSKLOEIDVFNQOLKEIRHSKQOLQOKORSLEA 626
581 LHKSSSEBELCQRLKEQDLALEKETASKLSEMDSFNNQLEKRSVYTOQLALEQHLKI 640
627 ARLKQKEQERSLE-LEKQK--EDQRRVQDERDKQMLEHVQOEQPPPKRPHEDRLK 682
641 KDKLKEIERKLEIQKQLEDEAARAKQKQKEMLRRESIRKEEE-----EKQRLQOE 694
663 EDSVKKAEERAKREMODKOSRLFHPQEPAKLATQAPWSTEGPLTISAQESVKVY 742
695 EKSQDKTQEBER--KAERKQSE-----TASH-----IYN 721
743 YRALYPPESRSHDEITIQGDIWVDESQTEGPGMLGELKQKGMFPYANAEKIPENEV 802
722 YRALYPPFARHNDKESFSSGDIQVDEKTVGPGMLYSGFGQKGFMCVNYEKLASSE- 780
803 PTPAKPVTDLTISAPKALARETPAPLPVTSSEBETNNMADBSSTPSSSNEKEPFDN 862
781 -----KALSPKRALPPTVLSLSTSTSS--SQPPASVTDYHNV--SFSNLTVNT-- 824
863 WDTMAQPSLTVPSAGQRLORSAPFPATATGSPSPVVGQEKVYGLQAOALYPMRAKD 922
825 --TW-----QKSAFRTVYSPG-SVSPFHQGGAVENLQALQALCSTAKKE 867
923 NNLNFKSDVITVLEEQDMMWFGVEYQOGKGFPSKYVLLISG-PYRKS--TSIDTPTES 979
868 NNLNFSKQDVTITVLEQENMMWFGVHGGRGMPKPSYVLLIGNEVQREBPALAAVATKK 927
980 PASLKRVASPAKPAIP--GEFIAMYYTESSEQGLDTQOQGDVYVYTKKQDMMTGVGD 1038
928 PTS--TAYPVYSTAYPVGEDYIALYSSVPEGLTFEGEELIIVTQKDESMWTGSIGE 984
1039 KSGVPSYVYVLEKQESGSGTAKGSLKKEPIAVIVASVYATGEBQTLAPGOLILRK 1098
985 RTGLFPSYVYVYVLEKQENFGANSKSGANSKKEPIAVISAVASGEBQSLAPGOLILRK 1044
1099 KNPQGMWEGELQARCKRQIGWFPYANYVYLLSPGTSKITPTLPTLPTAQAQVAVQVGYMD 1158
1045 KNTSGMWQEGELQARCKRQIGWFPYANYVYLLSPGTSKITPTLPTLPTAQAQVAVQVGYMD 1097
1159 YTAQNDDELAISKQOIIVNLKEDPDKMGVSGQVGLFPSNYVYVYLLTDDMDPSQ 1213
1098 YMANNEDELNFSKQOIIVNMKDDPDMMWQGETNGITGLFPSNYVYVYVYLLTDDMDPSQ 1152

```

```

RESULT 14
AA771160
ID AA771160 standard; Protein: 1683 AA.
XX
XX AA771160;
XX
XX 08-SEP-2000 (first entry)
XX
DE Rat phosphodiesterase interacting protein, M14.
XX
XX Rat; phosphodiesterase interacting protein, M14; PDE; cAMP-PDE;
XX cyclic adenosine monophosphate phosphodiesterase; antiaesthatic; asthma;
XX antiinflammatory; antiprostaglandin synthetase; antidiabetic; shock;
XX analgesic; immunosuppressive; antitumor; vasodilator; antihypertensive;
XX antidiabetic; urticaria; antiallergic; antiatherosclerotic; diagnosis;
XX antineoplastic; treatment; inflammatory disease; psoriasis; arthritis;
XX atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
XX eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
XX reperfusion injury; atopic dermatitis; diabetes insipidus;
XX conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
XX arterial restenosis; ankylosing spondylitis; transplant rejection;
XX graft versus host disease.
XX
XX Rattus sp.
XX
XX WO200027661-A1.
XX
XX 18-MAY-2000.
XX
XX 12-NOV-1999; 99WO-US26860.
XX
XX 12-NOV-1998; 98US-0108255.
XX
XX (STRD) UNIV LEIAND STANFORD JUNIOR.
XX
XX Conti M, Pahlke G;
XX
XX WPI: 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
XX polypeptide, useful for diagnosis and treatment of asthma, cystic
XX fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
XX
XX PS Disclosure; Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
XX M14 from rat. The protein module has the functions and properties of PDEs,
XX specifically cAMP-PDEs, and also targets them to specific subcellular
XX compartments. The present sequence
XX can be used in the diagnosis and treatment of disease conditions
XX associated with PDE activity. The diseases include asthma, cystic
XX fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
XX granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
XX septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
XX inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
XX distress syndrome, diabetes insipidus, allergic rhinitis, allergic
XX conjunctivitis, vernal conjunctivitis, arterial restenosis,
XX atherosclerosis, inflammatory diseases associated with irritation and
XX pain, rheumatoid arthritis, ankylosing spondylitis, transplant
XX rejection and graft versus host disease, disease conditions associated
XX with hypersecretion of gastric acid, and disease conditions in which
XX cytokines are mediators.
XX
XX Sequence 1683 AA;
XX
XX Query Match 48.6%; Score 3046; DB 21; Length 1683;
XX Best Local Similarity 50.8%; Pred. No. 1,1e-186;
XX Matches 652; Conservative 187; Mismatches 267; Indels 178; Gaps 33;
XX
XX 1 MAQFPFPGSLDVAITVEERAKHQDFLSLKPFIAGFTIGDQARNFFQSGLPQVLAQ 60
XX 2 MAQFPFPGSLDVAITVEERAKHQDFLSLKPFIAGFTIGDQARNFFQSGLPQVLAQ 61

```


Qy	61	IWALADNMNDGRMDQVEFSIAMKLIKLOQYOLPSTLPVVMKQOPVAISSAPAFGIGI	120
Db	61	IWALADNMNDGRMDQVEFSIAMKLIKLOQYOLPSTLPVVMKQOPVAISSAPAFGIGI	120
Qy	121	ASMPPLTAVAPVPMGSIIPVGMSPPLVSSVPAAVPLANGAPVIOPLPAFAHPAATWP	180
Db	121	ASMPPLTAVAPVPMGSIIPVGMSPPLVSSVPAAVPLANGAPVIOPLPAFAHPAATWP	180
Qy	181	KSSSFBSGPGSGLNTXKQKASFDVASAPPAEWAHPQSSRLKYROLFNSHDKTMSGHL	240
Db	181	KSSSFBSGPGSGLNTXKQKASFDVASAPPAEWAHPQSSRLKYROLFNSHDKTMSGHL	240
Qy	241	TGPQARTILMOSSLPOAOLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP	300
Db	241	TGPQARTILMOSSLPOAOLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLPVLP	300
Qy	301	PEYIPPSFRVRVSGSGSVISSSVDOQLPEEPSSSEDOQ-EKKLPVTFEDKKRENER	359
Db	301	PEYIPPSFRVRVSGSGSVISSSVDOQLPEEPSSSEDOQ-EKKLPVTFEDKKRENER	359
Qy	360	GSVELEKROALLERQKQERLAOLERAEOERKEREQOERAKROLEROLEKOLE	419
Db	361	GNLELEKROALLERQKQERLAOLERAEOERKEREQOERAKROLEROLEKOLE	420
Qy	420	ROREBERREIERREPAKKELEEROROLEWERNRROELNORVKEOEGTVLAKARRKTLF	479
Db	421	ROREBERREIERREPAKKELEEROROLEWERNRROELNORVKEOEGTVLAKARRKTLF	480
Qy	480	ELEALNDKXHQLEGLQDIERCLATQROEIESTNKSRELIAETHTLOOQLOESQOMLGR	539
Db	481	ELEALNDKXHQLEGLQDIERCLATQROEIESTNKSRELIAETHTLOOQLOESQOMLGR	540
Qy	540	LIPEKQIISDQIKOYQONSIRHDSILTLKRALBAKELARQOLREOLDVEVERTRSKLOEI	599
Db	541	LIPEKQIISDQIKOYQONSIRHDSILTLKRALBAKELARQOLREOLDVEVERTRSKLOEI	600
Qy	600	DVFNNQLEKEIREIHSKOOLOKORSLEAARLKOEORKSLEL	641
Db	601	DIFNNQLEKEIREIHNKQOLOKOKSMEARLKOEORKEIKIEL	642

Search completed: December 4, 2003, 15:18:34
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:12:19 ; Search time 24 Seconds
(without alignments)
4860.525 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQFPTFGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFQSGLPQVLAQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5063.5	80.8	1270	2 T09194	adaptor protein in
2	1735	27.7	1011	2 T13055	dynamitin associated
3	1698.5	27.1	1094	2 T13053	dynamitin associated
4	1333	21.3	1097	2 T31504	hypothetical prote
5	506.5	8.1	751	2 T34490	hypothetical prote
6	486	7.8	897	2 A54696	EGF receptor subu
7	485.5	7.7	896	2 S43074	epidermal growth f
8	410.5	6.5	1480	2 S48440	poly(A)-specific r
9	403	6.4	1381	1 S45781	probable calcium-b
10	391.5	6.2	1407	1 S28589	trichomyalin - rab
11	383.5	6.1	1909	2 A45592	liver stage antige
12	381	6.1	1898	1 A45973	trichomyalin - hum
13	375.5	6.0	1549	1 A40691	trichomyalin - she
14	359.5	5.7	737	2 T15597	hypothetical prote
15	359.5	5.7	1017	2 T15598	hypothetical prote
16	342	5.5	2526	2 T20531	hypothetical prote
17	342	5.5	2722	2 T20532	hypothetical prote
18	340	5.4	2738	2 E88320	protein P07A1.6
19	331.5	5.3	1069	2 T00377	KIAA0642 protein -
20	330	5.3	793	1 JH0628	caldesmon - human
21	327	5.2	1794	2 T84359	hypothetical diver
22	324.5	5.2	1181	2 C86349	P8K7.4 protein - A
23	323.5	5.2	408	2 T42650	hypothetical prote
24	321	5.1	665	2 JCT191	85K c-CD1-interact
25	319.5	5.1	1027	2 T46481	hypothetical prote
26	318	5.1	699	2 E84565	hypothetical prote
27	316.5	5.0	816	2 T17257	hypothetical prote
28	314.5	5.0	887	2 G88484	protein F23F12.8
29	309	4.9	4574	2 G02520	plectin - human

30	308.5	4.9	1701	2 T09127	probable erythrocy
31	307.5	4.9	1110	2 I51116	NF-180 - sea lamp
32	306.5	4.9	771	1 A33430	h-caldesmon - chic
33	304.5	4.9	585	1 A2168	involucrin - human
34	304.5	4.9	958	2 S47179	hypothetical prote
35	302	4.8	1052	1 A44937	kinetoplast-associ
36	302	4.8	4684	2 A59404	plectin (imported)
37	299	4.8	1233	2 T10989	serine/threonine p
38	297	4.7	2101	2 A42184	nuclear mitotic ap
39	296.5	4.7	3488	2 T34418	hypothetical prote
40	296.5	4.7	4957	2 T03455	ALR protein - huma
41	296.5	4.7	5262	2 T03454	ALR protein - huma
42	296	4.7	4687	1 A39638	plectin - rat
43	295	4.7	729	2 T50989	hypothetical prote
44	294	4.7	746	2 T47237	myosin II heavy ch
45	293	4.7	2442	2 T08621	centrosome associa

ALIGNMENTS

RESULT 1

T09194
adaptor protein intersecutin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T09194
R:Ramadhai, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castagnoli, L.; Cesareni, J. Biol. Chem. 273, 31401-31407, 1998
A:Title: Intersecutin, a novel adaptor protein with two eps15 homology and five src homol
A:Reference number: Z16605; MUID:99030416; PMID:9813051
A:Accession: T09194
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1270 <YAM>
A:Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AAC73068.1; PID:g2642625
A:Experimental source: cell type oocyte
C:Function:
A:Description: involved in endocytosis
C:Keywords: endocytosis

Query Match	80.8%; Score 5063.5; DB 2; Length 1270;
Best Local Similarity	80.6%; Pred. No. 2.7e+198;
Matches	980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;
QY	1 MAQFPTFGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFQSGLPQVLAQ 60
DB	1 MAQFPTFGSLDVAITVEERAKHDQFLSKPTAGFTGDQANFFQSGLPQVLAQ 60
QY	61 IVALADNMNDGRMOVERSIAMKLIKLTQGYQLPSTIPVWKQOPVAISSA--PAGIG 118
DB	61 IVALADNMNDGRMOVERSIAMKLIKLTQGYQLPSTIPVWKQOPVAISSA--PAGIG 118
QY	119 GIASWPPITAAVAVVMSGIPVVGMSPLVSSVPAAPVPLANGAPVPIQPIPAFHPAAT 178
DB	121 GIVGIPPLAAVAVVMSGIPVVGMSPLVSSV--TVPLNSGAPAVIQSPAPFAH-SAT 177
QY	179 WPKSSSFSSRGSGQLNTKLOKQSFVDAAPPAEWAAPVSSRLKQRLFNSHDKTWSG 238
DB	178 LPKSSSFSSRGSGQLNTKLOKQSFVDAAPVAVVAVPSSRLKQRLFNSDQKTWSG 237
QY	239 HLTGPQARTIIMOSLTPAQALASTIWNLSIDIDQDKLTAEFTLMLHLDVAMSGQPLPV 298
DB	238 NLTGPQARTIIMOSLTPAQALASTIWNLSIDIDQDKLTAEFTLMLHLDVAMSGQPLPV 297
QY	299 LPPEYIPSPFRVSSGSMVSISSSVSDQRLPEPSSSEBQOPKKLPVTEDEKRENF 358
DB	298 LPPEYIPSPFRVSSGSLTSSSVSDQRLPEPSEEBEPQNAKKLPVTEDEKRENF 357
QY	359 RGSVLEKRRQALLBQKREKERLAQLERABQEKREKREKROEAKRQLEKQEL 418
DB	358 RGNLELEKRRQALLBQKREKERLAQLERABQEKREKREKROEAKRQLEKQEL 417

QY 419 ROEEERERKRIERREAKRELEROROLEMERNRROELNORNEOGSTVLKAKRRTLE 478
 Db 418 ERQEEERKRRIERREAKRELEROROLEMERNRROELNORNEOGSTVLKAKRRTLE 477
 QY 479 FELELANDKKHQBEGKLQDIDRCRLATOROEIESTTNKSHELRIAEITHLQOOLQOESQOMLG 538
 Db 478 FELELANDKKHQBEGKLQDIDRCRLATOROEIESTTNKSHELRIAEITHLQOOLQOESQOMLG 537
 QY 539 RLIRPEKQLLSQOLKQVOQNSLHRDLSLTLKALAEKELARQOLREOLDEVERETRSKLOE 598
 Db 538 KMEPEKSLIIOQLKQVOQNSLHRDLSLTLKALAEKELARQOLREOLDEVERETRSKLOE 597
 QY 599 IDVFNNOLKELEIRHSKQOLKORSLEAARLKOROEKESLELKOKEDAORROEBK 658
 Db 598 IDVFNNOLKELEIRHSKQOLKORSLEAARLKOROEKESLELKOKEDAORROEBK 657
 QY 659 WLEHVQOEOPRPRKHEDERLKREDSYRKKEAEERAKPEMDOKOSRLFHPHOEPAKLAT 718
 Db 658 MODRKQEE--RYKFODEBEKEKEBESYQKOEVEK--KPEIQEKPNKPFHQPREPKLGG 713
 QY 719 QAPMSTKEGULTLSAQESVKKVYTRALYPRESNSHDEITIQPDYVWDESQGEFGML 778
 Db 714 QIPMNTEKAPLTLIN-QGDVKKVYTRALYPEDASSHDEITIEPDILINWVESQGEFGML 772
 QY 779 GGELKGGKGTGWPANAAEKIPENEVPTPAKPTDITLSAPAPLALREPPAPL-PTSSSPS 837
 Db 773 GGELKGGKGTGWPANAAEKIPENEVPTPAKPTDITLSAPAPLALREPPAPL-PTSSSPS 828
 QY 838 TTPNNMADPSSSTWPESSNEKEPETDMDTMAAOPSLTVPSAGOLKORSAFTPATATGSSPS 897
 Db 829 TNSNNMADPSSSTWPESSNEKEPETDMDTMAAOPSLTVPSAGOLKORSAFTPATATGSSPS 888
 QY 898 PVLGGKEVEGLQAOALYPMRAKKNHLENKSDVITVLEQODMMWFGEVQGGKWPFPKS 957
 Db 889 PVLGGKEVEGLQAOALYPMRAKKNHLENKSDVITVLEQODMMWFGEVQGGKWPFPKS 948
 QY 958 YVKLISGVRKSTSIDTQPTSPASLKRVASPAKPAIRGEELFAMTYBESBOGDLTFQ 1017
 Db 949 YVKLISGVRKSTSIDTQPTSPASLKRVASPAKPAIRGEELFAMTYBESBOGDLTFQ 1008
 QY 1018 QGDVIVMKKQODMMWTGTVGDKSGVFPNSNYRLKDSQESGAGTGSIGKKPEIAOVYAS 1077
 Db 1009 QGDVIVMKKQODMMWTGTVGDKSGVFPNSNYRLKDSQESGAGTGSIGKKPEIAOVYAS 1068
 QY 1078 YAATPEQTLTAPGQLILIRKKNPGMWEGELQARGGKQIGWPPANYVKLLSPGTSKIT 1137
 Db 1069 YAATPEQTLTAPGQLILIRKKNPGMWEGELQARGGKQIGWPPANYVKLLSPGTSKIT 1128
 QY 1138 PTELPRKTAQVAVQCVIGMYDTYANDELAFSKQIINLVANKEDPMWKEVSGQVGLF 1197
 Db 1129 PTELPRKTAQVAVQCVIGMYDTYANDELAFSKQIINLVANKEDPMWKEVSGQVGLF 1188
 QY 1198 PSNYVKLTITDMDPSQO 1213
 Db 1189 PSNYVKLTITDMDPSQO 1204
 RESULT 2
 T13055
 dynamn associated protein isoform 160-2 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13055
 R:Roos, J., Kelly, R.B.
 J. Biol. Chem. 273, 19108-19119, 1998
 A:Title: Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing protein
 A:Reference number: Z17594; MUID:98334647; PMID:9668096
 A:Accession: T13055
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1011 <ROO>
 A:Cross-references: EMBL:AF054612; NID:g2996029; PID:g2996030; PIDN:AAC39139.1
 C:Genetics:

QY	Query Match	27.7%	Score 1735;	DB 2;	Length 1011;
QY	Best Local Similarity	33.4%	Pred. No. 2.5e-63;		
Db	Matches 413;	Conservative 202;	Mismatches 332;	Indels 288;	Gaps 31
QY	11 SLDWVAIVERRAHNDQFLSKPLAGITTDQANRFFQSGLPQVLAQIYALADMMND				70
Db	4 AYDAAVAVTPRRRLTKQEQFRLQPOAGVVTQAQAKGFLQSLPPLILGQIYALADTBSD				63
QY	71 GRMDQVFSIAMKLTKLQGYOLEPSTLPYMKQOPVAISSAPAFGIGIASMP--LT				127
Db	64 GKMINERSIACKLINLTKLRGMDVPKVLPSL-----LSSP-----TGDPVSMTPRGSTS				113
QY	128 AVAPPMKSTIVNGMSPPLVSSVPYPAAVPPLANGAPPIYQPLPAHAPAAITPKSSPFR				187
Db	114 SLSPD-----PLKGI-VPAAVPVPPVPAVPA--VATVISP-PGVSVSPGPTPTPSN---				162
QY	188 SGPGSOLWTKQKQKOSFPVAPASAPAAEMAPQSSRLKTRQLFNSHDKTMISGHLTSPQART				247
Db	163 --PSPRHSTISERRAPSTISVN---QGEAAVQAQAKRKTQVFPNADRIRSGVLTSGSQARG				217
QY	248 ILMQSLPQAOALASIMWLSIDIDQDGLTAAEFILAMHLIDVAMSGQPLPVLPPYPIPS				307
Db	218 VLVQSKLPQVTLAQWLTSLDIDGRLNCDEFIAMFLFCERAMAGEKITPVTLPOEWVPPN				277
QY	308 FRVRSSGSMGVISSSVDQRLPPEPSSD-----EQPEKKLP--VYFEDKK				353
Db	278 LRKIKSRPG-----SVSGVSVSRPGSQPASRHHASVSSQSGVGVDADPTAGLPQGTSPEDKR				333
QY	354 RENFERGSVELLEKRRQALLQEQRKEQERLAOLERAQERKERERQEOEAKRQLELEKOLE				413
Db	334 KENYKQGAELDRKRKIMEDQQRKERERERERERERERERERERERERERERERERERERER				393
QY	414 KQELERQREBERKEKIRERPAKRELERQROLEMERNRROELLNQNRKEQGVTVLKAR				473
Db	394 KQELERQREBERKEKIRERPAKRELERQROLEMERNRROELLNQNRKEQGVTVLKAR				453
QY	474 RKTLEFELANDKHOLEGKLODTRCLATQROGIESNTNSRELRIEITHLQOOLQOS				533
Db	454 NTQLNVELSTINEKIKELISQRICTDRAGVTNVKTVIYIDGRTORDTSDMSMSQKARIKIQ				513
QY	534 QOMLGRLLIPER-----QLISDOLKQVQNSLHRDSILTLKRALFAELARQOLR				582
Db	514 NAKLILQLQLEBAKWEAKSAGSAGALGSMNAQEQ-----LNAAPAHQILLIINOJK				563
QY	563 EQLDEVEREPTSKLOEIDVFNNOLEKLR-----IHSQOOLQKQSLLEAKRLKQEQERK				637
Db	564 KDVENISKEIISKEDIDINTDVQMSLEKAEALATLKCEIDLKEDYDQVSTSLTELKYNKK				623
QY	638 SLELEKQKEDQQRQERDKQMLHNVQOEOBQRPKPHBEDRLKREDSYRKKAEABERAP				697
Db	624 NETSVSSAMWT-----GSSSAM-----BETGTTVTDPR-----AASNDISALAP				664
QY	698 EMODKQSLRFPHDEPAKLATQAPWSTTEKGQPLTISAQSVKVVYRYALYPRESSHDEI				757
Db	665 AVD-----LGGPAP-----EG-----FVKQYAVYENANNAABEI				693
QY	758 TIQPEDIMVWDBSQGTPEGMVGGELKGGKTGMFPNAPYAEKIPENEBVPTPAKPYTDLTSAPA				817
Db	694 TFVPEDIIILVPLEQNAEPGMLAGEINGHTGMPEESYVKLELVEGV-----A				739
QY	818 PKLALRETPAPLPLVYSSSPSTTPNNMADPSSSTWPSSSNEKPEITDMMDTWAAQPSLTPVSA				877
Db	740 PVAAY-LAPVDAQVADT-----YNDINT-----SITPA				768
QY	878 QQLRQSAFTPATATGSSSPPYLGGEGKEVGLQAGALYFWRAKKDNHLNFKNSDYITYLE				937
Db	769 S-----ADULTAG-----DV-----				778
QY	938 QODMMWFEEVQGGQKMFPSKYVKKLISGVYRKSSTIDTGPTEBPALKRVASPAAKPAIYG				997

```

Db      779 ----- 778
Qy      998 EEFIAMTYEESGDLTFQOQDVIVTKKQDMMTGTVGDKSGVPPSNVYRLKDSGSG 1057
      779 EYIYAIAPYESABEGDLSFSGEMWVVIKKEGEMWITIGSRGMPSPSNVYQADVGTAS 838
Qy      1058 YAG-----KTGSLGKKEPIAQNIVASVYATGPEQTLAPGQILIRKKNPGGWMGELQ 1110
      839 TAAAEVPSLDOGMRAKSEIAQVIAPEATSTEQSLTRGQILIMIRKKTDSGWMGELQ 898
Db      1111 AGCKRQIOWMPFANYKLLSPG--TSKITPTELPTKAVQPAVC-QVIGMTDTYAQNDD 1167
      899 AGRRRQIOWMPFANYKLLSPG--TSKITPTELPTKAVQPAVC-QVIGMTDTYAQNDD 958
Qy      1168 AFSKQIINVLNKEDPDMWKGVSQGVLPSPSNV 1202
      959 SFDKDIISVLGRDEBEMWKGELNGSLFSPSNV 993

```

RESULT 3

T13053

dynamlin associated protein isoform Dap160-1 - fruit fly (*Drosophila melanogaster*)C1:Species: *Drosophila melanogaster*

C1:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C1:Accession: T13053

J. Roos, J. Kelly, R.B.

J. Biol. Chem. 273, 19108-19119, 1998

A1:Title: Dap160, a neutral-specific Eps15 homology and multiple SH3 domain-containing pro

A1:Reference number: Z17594; MID:19834647; PMID:9668096

A1:Accession: T13053

A1:Status: preliminary; translated from GB/EMBL/DBJ

A1:Molecule type: mRNA

A1:Residues: 1-1094 <R00>

A1:Cross-references: EMBL:AF053957; NID:92984714; PID:92984715; PIDN:AAC39138.1

C1:Genetics:

A1:Gene: Dap160

A1:Cross-references: FlyBase:FBgn0023188

Query Match 27.1%; Score 1698.5; DB 2; Length 1094;
 Best Local Similarity 31.6%; Pred. No. 8.2e-62;
 Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;

```

Qy      11 SLDVAITVEERAKHQDFLSLKPIAGFITGDAARNFQSGLPQVLAQIWMALADMMND 70
      4 AYDAMAVTFRERLKYQEGFRALQPOGCVTGAQAKFFIQSQSLPILIQIWMALADTSD 63
Db      71 GRMDQVESIAMKLILKLGQYQLPSTLPVPMKQQPAVIAISAPAFGIGIASMP---LT 127
      64 GKNNINSEFIACKLIMIKLGMQVPCVLPSPSL-----LSSL-----TGDVPSMTFPGSTS 113
Qy      128 AVAPVPMGSIPIVVGMSPLVSSVPPAAVPPPLANGAPVYQPLPAFAHPAATMPKSSFSR 187
      114 SLSPLD---PLKGI-VPAVAVPVVAPVPA--VATVISP-PGVSVPSPGPTPTSN--- 162
Db      188 SGFGSOLNTKLQKQAFDVASAPPALEMAVPOSSRLKROLFNSHKTMSGHLTGQART 247
      163 --PSSHTSISERAPSIESVN--QGEWAVQAQKKTQVFRANRRTSGYITGSAQG 217
Qy      248 IIMQSSLPQAQLASIWNLSDIDQGLTAEEFLIAMLIDVAMSGQLPPLVPEYIPPS 307
      218 VIVQSLPQVTLAQITLSDIDQGLNDEFLIAMLCEKMAAGSEKIPVTLQPEWVPPN 277
Db      308 FRRVRSQSGMSVSSSVQRLPEPSSD-----EQPEKLP--VTREDK 353
      278 LKTKISRPG---SVSGVVRPSGQPSASRHASVSSQGVVADADTAGLPQGTSPEDKR 333
Qy      354 RENFERGSYLEKRGQALLQOQRQOEURLAQLERAQERKERERQOQANROLTEKOLE 413
      334 KENYVVGQALDPRKRIMEDQQRKEBERKERERKREADRERKRLAERKQOELELRQ 393
Qy      414 KORELERQEBERKEIEREAKRELEROROLEWERNRROELINRKNKEOEGTVLAKAR 473
      394 ROREIMERKEQKRELKAERKELERQOQEWQOARIAMWNAKEREOEVRVLKQAK 453
Db

```

```

Qy      474 RKTLEFELEALNDKKHOLEGKLDIRCRLATOROIEESTNKSRELIATITLQOOLQES 533
      454 NQULVELSTLNKELKELSORICDRAGVTANKYVTDGKTRDRDSMSMSQKARIKQ 513
Db      534 QOMLGRILPEK-----QILSDQLQVQNSLHRSDLLTKRALAEKELARQQLR 582
      514 NAKLLQTLQERAKWEAKSASGALGGENAQOE-----LNAFAHQIILINQJK 563
Qy      583 EQLDEVERETRKLOEIDVFNQKELRE-----IHSKOQLOKQSLAERAKKQOEK 637
      564 DKVENISKEIESKEDINNDVQMSLEKALSALTTCEDILKEVDVORTSVLELKYNK 623
Qy      638 SLELEKQKDAQRVOERDKQWLEHVOEEOGRPKRPHEDRLKQEDSVKKEAEERAKP 697
      624 NETSVSAMDT-----GSSSA--BETGTYTDPY-----AVASDISALAP 664
Db      698 EMQDKQSRLEFHPQEPAKLATQAPWSTTEKGPLTISAQSVKVVYRYALPPESRSHDEI 757
      665 AYD-----LGSPAP-----EG-----FVKYQAVYEFNANNAEEL 693
Qy      758 TIQPGDIWVWDSQTEGREGWLGELKGTGMPFANYAEKIPENYPTPAKPVTLDTLSDA 817
      694 TEVPEDILIVPLEQNAEFGWLAGETNGTGMFPESYVEKLVEGEV-----A 739
Qy      818 PTLALREPPAPLPVTSSESTTPNNMADPSSSTWSPSSNEKPTDMMDTMAQPSLTVPSA 877
      740 PAAAN-EPADVDAQVDT-----YNDINNT-----SSTIPA 768
Db      878 GOLRQSAFTPATATGSSPSPLGQGEKEVGLQAQALYPMRAKKNHNLFNKSDVITVLE 937
      769 S-----ADLTAG-----DY----- 778
Qy      938 QQDMMWFGEVQCKMFPKSYVKLISGPVKSTSIDTGPSTPACLKRVASPAKAPALPG 997
      779 ----- 778
Qy      998 EEFIAMTYEESGDLTFQOQDVIVTKKQDMMTGTVGDKSGVPPSNVYRLKDSGSG 1057
      779 EYIYAIAPYESABEGDLSFSGEMWVVIKKEGEMWITIGSRGMPSPSNVYQADVGTAS 838
Qy      1053 -----SESG----- 1057
      839 TAAAEVPSLDOETLNGAATAPAVEAQEVQPLPVQEPSSEQPISSPGAGABAHND 898
Db      899 IDTEVSQINTQSKTOSSEPAESYSRPMSTSSMTPGMAKRSIEIAQVIAPEATSTEQLS 958
Qy      1058 -----TAGKT-----GSLGKKEPIAQNIVASVYATGPEQTL 1087
      1088 LAPGQILIRKKNPGGWMGELQARGKKRQIQGMPFANYKLLSPG--TSKITPTELPTKA 1145
Db      959 LTRGQILIMIRKKTDSGWMGELQARGRRQIQGMPFANYKLVQGGNSGRNTPVSGSRLE 1018
Qy      1146 VQPAVC-QVIGMYDTAQNDELAFSAKQIINVLKKEPDMWKGVSQGVLPSPSNV 1202
      1019 MTEQILDVIAIYPIKAQNDELSTFDKDIISVLGRDEBEMWKGELNGSLFSPSNV 1076
Db

```

RESULT 4

T13104

hypothetical protein Y116A8C.36 - *Caenorhabditis elegans*C1:Species: *Caenorhabditis elegans*

C1:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C1:Accession: T13104

R. McMurtry, A. submitted to the EMBL Data Library, October 1999

A1:Reference number: Z21041

A1:Accession: T13104

A1:Status: preliminary; translated from GB/EMBL/DBJ

A1:Molecule type: DNA

A1:Residues: 1-1097 <W11>

A1:Cross-references: EMBL:AL117204; PIDN:CAB55138.1; CESP:Y116A8C.36

A1:Experimental source: clone Y116A8C

C:Genetics:

A:Gene: CESP:X116A8C.36
A:Introns: 50/3; 245/3; 411/3; 486/3; 697/1; 820/3; 926/2

Query Match 21.3%; Score 1333; DB 2; Length 1097;

Best Local Similarity 29.4%; Pred. No. 5,6e-47;

Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

```

QY 15 WAITVERAKHDQQLSLKPIAGFTTQD-----AANFFQSGLPQPVLAQIWMALDM 67
DB 5 WEVSAAEQKPFAMF-----GQLTGQPFMDAVATANALMRSLPTQVLSQIWMALSD 57
QY 68 NNDGMDQVSIAMKLIKLGQQLPSTLPVVKQOPVAISSAPFGISAMPPLT 127
DB 58 DKDGLDIREVISIMKRLALNCIAGIPPIPPQLLVPA--KNAFPWPGS----- 107
QY 128 AVAPVPMGSIPIVNGSPPV-----SSVPPAAVPLNAGAPVPIQPLPAFAPAAATWPK 181
DB 108 -----RHGSVDYSQTLPPALIDRRMSQSYIPSA--FVSIAGTP----- 142
QY 182 SSSFSRSGSGSLNTKLOKAQSFVVASAPPAEAAVPOSSRLKTRQLFNSHDKTMSGLT 241
DB 143 SSRNHSISAGSPLN-----NDRNVEBGRQLEMAIIPHNNKLYKSQLFNALDKERTGLS 196
QY 242 GPQARTIMOSSLPQAOIASIWNLSIDIDQGLTAEBFLIMHLIDVAMSGQPLPVLP 301
DB 197 SQVGRSALGSLPNNVLNHLIFLSDVNDQGLSDVEAISYMLEMFKSGFALPKITPL 256
QY 302 EYIPSPFRVRVSGMSVSISSSVDRLPPEBPSSEDEQPEKLPVTEDEKRENFERGS 361
DB 257 ELV-----RMCGISSRANNTPELEBGAEPQ--KSPAKTFEDKQDVLKSGQ 303
QY 362 VELEKROALLBQKKEQERLAQLERAQEKERERQOEAKQOLEKQLEKQLEKQ 421
DB 304 AELERRRVLLEEBERRRAVEKKEBEBAKKNRKEKEREQAEVERQAELEERQIIIAQ 363
QY 422 REEE---RRKEIERR-----EAAKRELE---ROQLMEMNRROELNORNEQEGTVV 469
DB 364 REEBEKRRLEMERREDEDEKRRKQMEKAKKQOVQMPCKKNPFNOKQOEHERLAQ 423
QY 470 LKARRKLEFLEALNDKQKLEGKLODIRCLATQROEIBSTNSRELIATIEITHLQOQ 529
DB 424 RQQRKTLQFQALDEKVIDEVDIGAKAEVAEVGTGFIEMRSTREKVARL-----KE 479
QY 530 LOESQOMGLRILPEKQIISDQKQVQNSLRHDSILITLKALEAKELARQOLRELD-- 586
DB 480 LQETNQ---KTAIESQELGHQ--LQOKSAHKE--TORKSELEALRRKRDRIKKAIEDPA 533
QY 587 -EVEKETSQKQEIIVF--NNOLKELREIHSKQQLQKQSLBAARLKQKOEKSELEKQ 644
DB 534 LELSTEKEKSYNQTEILKTNKEKTKTDVYSK-----LVAKKEEYRNSFEL-- 578
QY 645 KEDAQRVQERDKQMLEHVQOEQEPRPKP-----HEEDRLKREDSVRK 688
DB 579 ----LVHAQTHARSKIGEFKAKSAPASAPAPAPATTNNGPAPANNDAFGEED--K 631
QY 689 KEAERARAPENODKOSRLFPHQEPAKLATOAPMSTTEKGPITISQSSVKVYTRALYP 748
DB 632 TDAQRFADPGATST-----ADPFAQIACAP--AHSKGAVDQSAFNIHDTYKCRALFA 683
QY 749 FESRSHDEITTOPGDIWVWDSQTEPGMIGELGKGTGMPANVAEKIPENEVTPAKP 808
DB 664 FEAREDELSFEPGVITIVFQSHAAEPGRAGQLEKQGMPEAAVEAIA--AVTPG-- 739
QY 809 VTDLTSAPAKLALRETPAPLPVTSSEPTTPNNNADFSSTWSSSNKEPETDNMTAA 868
DB 740 -----GDPLIQMPPNMTWSSVDQIGV--KAARKAEI-----AA 772
QY 869 QPSLTVPSAGQLQKRSATPATATGSSPSPVGCGEKVEGLQALYIMPRAKGNHILFN 928
DB 773 AMGLTEGA-----PPASSAPAAAAYIS-----QCIQFQWRANNEDELSFA 814
QY 929 KSDVITVLEQDMMWFG--EVQGGKGMFKPSYKLTSGPVKSTSLDTGPTSPALKRVA 987

```

```

DB 815 KGDITVLEKQEMKMKRNAPAGEIGWFKSVKVGATTTSTTPI--VSPSASGAPGAA 873
QY 988 SPAKPAIPGEE-----FIAMTYESSBGDLTFQGGDYIVYTKKQDWMWG 1034
DB 874 AGAQYDVAPSVLTLQASETAPQOQLYVIVYDFEAVETTLALHGDITLVLEKDEWKG 933
QY 1035 TVGDKSGVPPSNYRLDSESGTAGKT--GSLCKKEI--AQVIASVATGPEQTLA 1089
DB 934 KGNREGIFFPANVEISVQAGDPPTPTQAPTPAAPTVLCEAKVVDVPAASAPQGLK 993
QY 1090 PGQILIRKKNPGWMEGELQARCKRQIGWFPANVYKLSPGTSKITPTELPTAVQPA 1149
DB 994 VGEIVKIRKSAACWMEGELIRNGKP--IAGWPFPEYKVLBEAASPAT----- 1040
QY 1150 VCQVIGMYDTAQNDELAFSKQIIVLANKEDPMWK--EVSQGVGFPSNYVK 1203
DB 1041 --RATAVYDEASQPDLEIGFTGDIIVTDKSEAWMSGHRQDPSKGLFPSNYVQ 1095

```

RESULT 5

T34490
hypothetical protein ZK1248.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34490

R:latreille, P.
submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid ZK1248.

A:Reference number: 221534

A:Accession: T34490

A:Status: preliminary; translated from GB/EMBL/DBJ

A:molecule type: DNA

A:Residues: 1-751 <LAT>

A:Cross-references: EMBL:U29244; PIDN:IACT1084.1; GSPDB:GN00020; CESP:ZK1248.3

A:Experimental source: strain Bristol N2; clone ZK1248

C:Genetics:

A:Gene: CESP:ZK1248.3

A:Map position: 2

A:Introns: 37/1; 74/3; 118/3; 355/2; 661/1; 728/2

Query Match 8.1%; Score 506.5; DB 2; Length 751;
Best Local Similarity 24.7%; Pred. No. 1.3e-13;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

```

QY 15 WAITVERAKHDQQLSLKPIAGFTTGDQARNFFQSGLPQVLAQIWMALDNNDGRMD 74
DB 111 WPIPTQAKYDSIFQSLNPNVNGKLSGAHYRPVLMNSGLDAHALARIWEISDDQKGLD 170
QY 75 QVESIMKLIKLIKLGQQLPSTLP--VMKQOPVAISSAPFGIGIASMPILTAVAPV 132
DB 171 RIEMSVALLHLVYRSLQSDPVPAPQLPVNLHPSKMYAHSSNF-----AAPP--HPPR 221
QY 133 PM-----GSIYVVGSPPLVSSVPPAAVPPILANGAPVITQPLPAFAPPA----- 176
DB 222 PMGSSRAGSVTSLDDVMSQS--YSATMPRAVP-----QGRAYSAQAHNNGRTS 271
QY 177 -ATWPKSSSFSRSGSGSLNTKLOKAQSFVVASAPPAEAAVPOSSRLKTRQLFNSHDKT 235
DB 272 GASTPISASHS-----IHSF-----PAGEWPIITNGD--YADQPAQTDTN 308
QY 236 MSGHLTPQARTIMOSSLPQAOIASIWNLSIDIDQGLTAEBFLIMHLIDVAMSGQPL 295
DB 309 KDGLVDDGMDKAPMTTGLSAQILAHWALADIKKQQLNLEQALTMHLIDMKRGESEI 368
QY 296 PPLVPPPEYIPSPFRVRVSGMSVSISSSVDRLPPEBPSSEDEQPEKLPVTEDEKRE 355
DB 369 PSELPHILIPSPFRPPEPRLHHPAQSIVSTPQLEPATSWIEIKALB-----GE 417
QY 356 NFERGSVELEKRRQALLBQKKEQERLAQLERAQEKERERQOEAKQOLEKQLEKQ 415
DB 418 NEF--MKQLAESIQSMVLERTKTAEEAVIQLE--ADMTIKNSISIK-----NLQVELATLESTV 470

```

QY 416 RELEROREBERRK-----EIERREAAKELEROROLEMERNRROELLNQRNKEOGTVV 469
 Db 471 KQLEQKGAATRLADVDYQIEQLBSA-----CAQKETKEDTERKMQQIDEDAKNED--- 524
 QY 470 LKARKRTLEFELBALNDK- -HOLEGKLODIRCLATOROEIESTNKSLELAIETHIQ 527
 Db 525 CKA-----NDEKEMELKEIEMLDNQFVGEIVKETSQGEQVAELT--- 569
 QY 528 QOLOESQOMLGRILPEKQILSDOLKQVOQNSLHRDSLLTKALALEKELARQOLR- EQUD 586
 Db 570 -----TLERKE-ARDQIQMEKLD 586
 QY 587 -EVERETSQLOEIDVFNNOELKEIRHSGKOOLOKORSLEAARKKQEKERSLELEKQ 645
 Db 587 AAIENTTKLTREVSDAVKESSEEMQI-----LRSGRLSTVYIDSLSDTVIGETNG 641
 QY 646 EDAQRVRQERDKOMLEHVQOESQPRPRKPHEDRLKREDSVRKEAEBAKEMODKQSR 705
 Db 642 TSSQ-----NHVQQPPDP-----FASADANPAAD----- 665
 QY 706 LFHPHOEPKATQAAPWSTTEKGPLTISAQESVKVYVYALYPFESRSHDEITIQGDIV 765
 Db 666 -----PFAVDQ-----FGSSGH----- 678
 QY 766 MVDESQTEGPGWLGELGKGTGMFPAN--YAEKIPENEVTPAKPYTDLTSAAPRLALR 823
 Db 679 -FDAFPPIDPFAQCG-----FSDSGFAQSAAPK---PAPP-----RPAPEKAKR 719
 QY 824 ETPAPLPVTSS- PSTTPNNMADFS 847
 Db 720 ETPVNDPFAPSQGSTQFAGFADFA 744

RESULT 6
 AS4696
 BGF receptor substrate epais - mouse
 C.Species: Mus musculus (house mouse)
 C.Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 05-Nov-1999
 C.Accession: AS4696
 R.Fazoli, F.; Minichiello, L.; Matoskova, B.; Wong, W.T.; Di Fiore, P.P.
 Mol. Cell. Biol. 13, 5814-5828, 1993
 A.Title: epais, a novel tyrosine kinase substrate, exhibits transforming activity.
 A.Reference number: AS4696; MUID:93361014; PMID:7689153
 A.Accession: AS4696
 A.Status: preliminary
 A.Molecule type: mRNA
 A.Residues: 1-897 <FAZ>
 A.Cross-references: GB:121768; NID:G404756; PIDN:AAA02912.1; PID:G404757
 C.Superfamily: calmodulin repeat homology
 C.Keywords: EF hand; phosphoprotein
 F:48-80/Domain: calmodulin repeat homology <EP1>
 F:160-192/Domain: calmodulin repeat homology <EP2>
 F:223-255/Domain: calmodulin repeat homology <EP3>

Query Match 7.8%; Score 486; DB 2; Length 897;
 Best Local Similarity 21.3%; Pred. No. 1e-12;
 Matches 205; Conservative 188; Mismatches 298; Indels 322; Gaps 31;
 QY 15 WAITVERAKHQOQFLSLKPIAGFITGDQARNFFQSGLPQVLAQITMALADNMNDGRMD 74
 Db 122 MAVKSEDKAKYDAIFDSLSFPVDGFLSGDKVKPVLNLSKLPVEILGVMRLSDIDHDGKLD 181
 QY 75 QVEFSIAMKLKIKLQGYLPESTLPVWKQCPVALISSAPAFIGIGIASMPPLTAVAPVM 134
 Db 182 RDEFVAMFLVYCALE-----KEPVVM 203
 QY 135 GSIPVVGMSPLVSVVPAAPVPLANGAPVPIQPLPAFAHPATATWPKSSFSRSGPQSOL 194
 Db 204 -----SLPPLVPP-----SKR 215
 QY 195 NTKLOKAGSDVASAPPAEMAVPOSSRLKYQLFNSHDKTWSGHLTGPQARTILMQSSL 254
 Db 216 KT-----WVVSPEAKAKYDEIFLKTQKMDMGVYSGLEVETFLKTL 257

QY 255 PQOALASTWNLSIDIDQDKLTAEERILAMHLIDVAM- SGQPLPVLPPEYTPPSRRVRS 313
 Db 258 PSLALAHWISLCTDYKCGCKLSKQFALFHLINQKILGIDPPHSLTPEMPPS---DPS 314
 QY 314 GSGMVISVSSSVQDRLPEEPSSEDEQPEKKLPVPEDEKKEENFPGSVTELEKRAQALLE 373
 Db 315 SIQKNITGSSPV-----ADFSAIK-----ELDTLNEIYD 344
 QY 374 QQRKEORLAQLEAEQERKERERQEOBAKQLELEKOLEKQERLEORREBERREIERR 433
 Db 345 LQREK-----NNEQDLKEKE-----DTVQKRTSEVQ---DLQDEVQRE 380
 QY 434 EAAKELEROROLEMERNRROELLNQRNKEOGTVVLAARKRTLEFELBALNDKQOLBG 493
 Db 381 SINLOKLAQKQ-----QVOELLGE-----LDEQAKOLEE 410
 QY 494 KLQDIRCLATOROEIESTNKSRELRLAETHLQOOLQESQOMLGRILPEKQILSDOLKQ 553
 Db 411 QLOEVRKCKAEBAQLISSLK-----AEITSQESQISSYEBELK-----AREELSR 456
 QY 554 VQONSILHRDSLTLTKRALAEKELARQOLREQLDEVERE-----TRSKLOEIDVFNNOQKE 608
 Db 457 LQOETAQ-----LEESVSGKAQLEPLIQHLOESQOEISSQWRLEMKDLTDNNQ--- 507
 QY 609 LREHSGKOOLQKQ-----RSLAARLKQKQEKRSLELEKQKEDAQRYQ 653
 Db 508 -----SNWSSPSQSVLVNGATDYCSLSTSSSETNPFNHAEGQNNLIESBPHQESSVSS 562
 QY 654 ER-----DKOMLEHVQOERQPRPRKPHEDRLKREDSVRKEA-----EERAKPEMODK 702
 Db 563 PEIAPSDVTDESEAVTAVAGNEKVTFR--FDDDKHKEEDPRVNESSSLTDVAVDNLDLF 620
 QY 703 QSRLF-----HPQEPKATQAAPWSTTEKGPLTISAQESVKVYVY 743
 Db 621 QSDPFGSDPFKMDPFKIDPFQGPFPFGSDPFADCFQKQSTDPFTTSSSTDPSASNN 680
 QY 744 RALYFESRSHDETIOGDIVMVDESQTEG--GWLGEKLGKGTGMFPANYAEKIPENE 801
 Db 681 SNTSVEETWTKNNDPPAPGTVVAAASDSATDPFASVFENESFG--DGFADFSTLSKVNND 739
 QY 802 VTP-----AKPVTDLTSA-----PA--PKLARETPAPLPVTSSSEPSTNN 842
 Db 740 AFNPTISSSTSSVTIAKMLEETKASKSDVPALPPKGTPTPRPCPP-----PKRPLN 794
 QY 843 WADFSST-----WPSSNKEPFTDNDMTAAOPLSTVPSAGOLRQSAFTPATATGSS 895
 Db 795 KLDSSDPLKLNDDPFQPFPGNDSPEKXKDDMFC-----DPTFSSTTKKE 838
 QY 896 PSP 898
 Db 839 ADP 841

RESULT 7
 S43074
 epidermal growth factor receptor substrate - human
 C.Species: Homo sapiens (man)
 C.Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
 C.Accession: S43074; I38525
 R.Bernard, O.A.; Muchauffe, M.; Mecucci, C.; van den Berghe, H.; Berger, R.
 Oncogene 9, 1039-1045, 1994
 A.Title: A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4,
 A.Reference number: S43074; MUID:94181254; PMID:8134107
 A.Accession: S43074
 A.Status: preliminary
 A.Molecule type: mRNA
 A.Residues: 1-896 <BER>
 A.Cross-references: EMBL:Z29064; NID:G470034; PIDN:CAA62305.1; PID:G470035
 R.Wong, W.T.; Kraus, M.H.; Carlomagno, F.; Zelano, A.; Druck, T.; Croce, C.M.; Huebner,
 Oncogene 9, 1591-1597, 1994
 A.Title: The human epais gene, encoding a tyrosine kinase substrate, is conserved in evo
 A.Reference number: I38525; MUID:94239734; PMID:8183552


```

QY 161 GAPVIOF-----LPAPAPATWPKSSFSRSGPSOL-----194
Db 506 GA---LQSYTGCGYGVWPOQSPASM--PNLSFNOQLOSLTGLQPOPTGFLPPSNFS 560
QY 195 -----NTRLOKASFDVASAPPAEWAPOSSRLRYOLFNSHDKTWSGHLT 241
Db 561 ATMPILTAQKTGFONNEIYTKSNFNNNLINDSSQDKISTEKSFLFYIFEPDTONKGLD 620
QY 242 GQARTILMOSSLPOQOLASIMWLSIDODGKLTAEFFLAWHLIVAMSGOPLPVLP 301
Db 621 SPTAIEIFRKSGLINRADLEQIWNLCIINNVTGQINKOEFGALGMHLYVGLKNGKPIPVLP 680
QY 302 EYIP-----PSF-----RRVSG--314
Db 681 SLIPSTKLLDNKQKLTPTTTEKPSFGKIDALSTYKNNDDVLPTNRRKRYISANQ 740
QY 315 ---SGMSVSISSVDQRLPEEPSEDEQPEKKLPVTFEDK---KRENERGSELK 367
Db 741 EEOSSFSFSAKSVNH-----SSSTLQTDIDISVDKTEVKAKPKYAGFSR---EINK 791
QY 368 RQALLQCKRQER-----LAQLERAQERERQOEAKRQLE 407
Db 792 NIASLENEIKNISNPENCYDSSIPSDLTSPFALIKLNLNLEISTINNEITNAKIOLY 851
QY 408 LEKQLEKQRELEKQREERKEIERREAKRELEKQLEWERNRQELINQKNEQEGT 467
Db 852 RKK---NPSSTIGSGNGEITENDRKAKASRLLR-----ABMSALTGKSTESDS 899
QY 468 VVLKARKKLTLEFE-LEALNDKHKQLEGLQDRCRLATQROETSKRELRALAEITL 526
Db 900 LEMBEDQSAEIKRIQENGKQEI--IKDIRSSISDISASIKSMTOSNM---ISN- 951
QY 527 QOOLQSOQMLGRLLPEKQILSDLKQVQNSIHRDS-----563
Db 952 -GEFERMEFGIGLEDVREFFL-DOLKSNKNSVTESSPVPSSTPPVDRSSPSYSQF 1009
QY 564 ILTLKRALEAKELARQLEOLDEVERETRSKQLEIDVFN--NQLKELAEIHSKQOL- 618
Db 1010 KTAEEBAVYLKQAKQKMEKLAQPKNRNVTQSSRSISSENSRQPOQIAGSSNLVLP 1069
QY 619 ---QKRSLEBAARKQERKESLEKQK-----646
Db 1070 RAPTQREKYEVAQPTQVQSTQPVQPTQPVQPTQPVQPTQPVQPTQPVQPTQPVQNY 1129
QY 647 DAQRVQERDKQMLEHVQOEOQRPKPHBEDRLK-----EDSVKKEAE-693
Db 1130 NAKQESDEDEDEDEKRLQEBELKRLKLKKKADKEKRLALRKQIEDAQNESDEEETNGK 1189
QY 694 -----RAKP-----EMQDKQSLFPHHQBPA--KLATQAPMSTTEKGLPILTA 734
Db 1190 NFGHVNVPQAPVAPASAPFQSNSTNAPSVHAAVTPAAGKSTGLPSTTMGHNPFKCA 1249
QY 735 QESVKKVYRALYPFESRSHD-EITIQPDIVWVDESQGEPCMLGELKGTGWMPANY 793
Db 1250 SAS-----STSTPAPAAEMQRRQKQ-----LDDEE--DQMSDEDSNNR-----VAV 1292
QY 794 AEKIPENEVPTP---AKPVTDLTAPAPKALRETAP--LPVTSSEPTTPNNMADS 847
Db 1293 DNKVEBAKIGHPDHAPAPV---TAAPLPV---TPVPAPAVPQANTSNESKSPITPA 1345
QY 848 STWPSSNS-----KPTNDMTMAQPSLTVPASQALQRS--APAPATATGSSPVPV 901
Db 1346 PIPPSVTOQEPVPLAPLPVADQFQEPPLPSAPALATAVOKSGSSTPALAGVLP-1402
QY 902 QOEKVEGLQAOALYPMRAKDNHLENK-----SDVITVLEEQDMMWFGEVQOK 951
Db 1403 --PPLPTQOASTSEPIIAHVVDYNGAEKGTGAVGSDSDDIVLSIPS-----VGTDEEE 1455
QY 952 GMPFKSYVKLISGP 965
Db 1456 GAQPVSTAGIPSP 1469

```

```

RESULT 9
S45781
probable calcium-binding protein YBI047c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBI0520
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 19-Apr-2002
C:Accession: S45781; S50284; S45782; S39841; S37339; S42498
R:Goffeau, A.; Joniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45781
A:Accession: S45781
A:Molecule type: DNA
A:Residues: 1-961 <GDP>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBI047c
A:Experimental source: strain S288C
R:de Wergifosse, P.; Jacques, B.; Joniaux, J.L.; Purnelle, B.; Skala, J.; Goffeau, A.
Yeast 10, 1489-1496, 1994
A:Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome II
NA-binding protein.
A:Reference number: S50284; MUID:95176707; PMID:7871888
A:Accession: S50284
A:Molecule type: DNA
A:Residues: 1-961 <DEP>
A:Cross-references: EMBL:X78214
A:Experimental source: strain S288C
R:Dubois, B.; El Bakoury, M.; Giansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.,
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45782
A:Accession: S45782
A:Molecule type: DNA
A:Residues: 579-1381 <DUB>
A:Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBI047c
A:Experimental source: strain S288C
R:Scherens, B.; el Bakoury, M.; Vlerendeels, F.; Dubois, B.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye
A:Reference number: S39824; MUID:94205266; PMID:8154187
A:Accession: S39841
A:Molecule type: DNA
A:Residues: 579-1381 <SCH>
A:Cross-references: EMBL:Z23261; NID:9313733; PIDN:CAAB0797.1; PID:9313748
A:Experimental source: strain S288C
C:Genetics: SGD:EDI1; MIPS:YBI047c
A:Gene: SGD:EDI1; MIPS:YBI047c
A:Map position: 2L
A:Superfamily: Yeast probable calcium-binding protein YBI047c; calmodulin repeat homology
C:Keywords: calcium binding; EF hand; transmembrane protein
F:167-199/Domain: calmodulin repeat homology <EF1>
F:560-576/Domain: transmembrane #status predicted <TMM>

Query Match
Best local similarity 6.4%; Score 403; DB 1; Length 1381;
Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

QY 3 QEPYFP-GGS-----LDWAIIVEBAKHDOQLSLKPIAGITGQARNFFQSGLP 54
Db 109 QNPAPMOGSGATGNTNNTDIPALSSNDIAKESQLDRTAKGAQYAGDKAKDIFPKARLP 168
QY 55 QPVLQIYALADNMNDGMDQVEFSIAKMLKTLKXQGYQLPSTLPVWKQCVVAISSAPA 114
Db 169 NOTLGEIYMLCDRDSAGVLDSEFTMAWYLLQLCSHPSNMTTPAVLPQ-----219
QY 115 FQIGIASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLANGAPVPIQPLAPAF 174
Db 220 -----LMDSI-----RLPEVVVNGPNRTTPLSANT 245
QY 175 PAATPCKSSFSRSGPSQOLTKLOKQSFVASAPPAEWAPOSSRLRYOLFNSHDK 234
Db 246 GVSSILTRHSTISRLSTGAFSNA-----ASDWSLSPEKKQQFPAIDSLDK 290
QY 235 TMSGHLTPQARTILMOSSLPOQOLASIMWLSIDODGKLTAEFFLAWHLIVAMSGP 294

```

291 QHAGSLSSAVLVPFLLSRNLQETLATITMDLADHNNMFFTKLEBALMFLIKKNAVE 350
 295 LPPVLPPE-----YIPSFRRVRSGSGSVISSS----- 324
 351 LPDVIPENLQSPALGLYPNPLPQOQSGAPQAIIPSRASKPSLDQMPHQVAPAVNTQPT 410
 325 VDQRLPE-----EPG-----SEDEQP 341
 411 VPQVLPQNSNNGSLNDLLALNPSFSSPPTAQTVAQVNTNNTNSFYDNNNGOATLQOQOP 470
 342 EKLVPTEDEKKRENFERSVELEKRRQALLQCKEKEERL----- 382
 471 QQPPLTLSSSGLKFFTPTS-----NFGOSIIEKEEBEQLERESSDTFSQAQPPVPKIAS 526
 383 ----- 382
 527 SPVKRTASTTLTPQVNPFSVPMAGAAATSAATGAAGAAGAAAGASAFSSSNNNAEK 586
 383 -----AQLEBAEOKERERQEQEAKROLELEKQRELEKOREBERKEIER 432
 587 QDLFDGASAKQSLNATTEMANLSNQVNSLSKQASITNDKSRATQELKRVTEMKNSIQI 646
 433 REAAKRELEKROLEMEHNRROELNQNKQEGTVLVKARKTLEFELALNDKQOLE 492
 647 KLNMRSTHDQNVKQTEQLEAQLV-----QVKNENE-----TLAQOLAVSEANTYHAE 694
 493 GKLODIRCLATQROEISTNKSRELIAET-----THLQOQLOESQOMGLRIPEKQILS 548
 695 SKLNE-----LTTDQESQTKNAELKEQITNLNSMTASLQSQINLEKQ----- 737
 549 DOLKQVQNSLHRDLSLTKALEKELARQLEQLEVEBERTSKQLEIDVFENQKE 608
 738 -QVKN-----ERSNVQVNSKQLELNQVYVANLQKEIDQLG-----EKSIVYTKQKE 783
 609 LREIHSKQOLQKORSLEAARLKQEQE-----RKSLEY-----EKQEDAKRQVQERKQMLBHV 663
 784 LND-----YQKTVBEQHAQLOAKYQDLSNKTOTLDREKQLEBNRQIEQENLYHGHV 837
 664 QOEQOPRRKPHREDRLKREDSVKKEABERAK-----PEWQKQSLFHPQEP-- 713
 838 SKLQK-----MFDLSQKAKSFEKADQELKERNIEYANNVNELESEKQML--AMQLEPD 890
 714 -----AKIATQAPMT-----TEKGPL-----TIS-----AESVAVVYVRYRLLYPRESHOI 757
 891 AKDIITAKSASNTDTTKKATSRKGNVHEDTVSKFVETTVENSNNLVNRYKDEBKTERTES 950
 758 TIQPGDI--VMVDESQTEGFGMLGELKGTGMPFANYAEKIPE-----NEVPTP--AKPV 809
 951 DVFPDRDVTLSQDSQSEANANTNGQSGNETANP--NLNETLSDRDGDLDNEIGIPRSQSL 1009
 810 TDLTSAPKALARETPAPRLPYTSEPTTPNNMADFSSSTWSSNEKEPETDNDQWMAQ 869
 1010 TSSVANNAPQ--SVRD--DVELPETLEERDTI--NNTANRONT-----GNLSHIDGEMEATPAT 1062
 870 PSLTPASGQLRQBAFAPATATGSSPSPVLG--QGEVYEGLOAQ--ALYPMWAKKXON-- 923
 1063 ASTDVL-----NETTEVIEDGSTITKANSNEGESVSIQESPKISAQ--KAKTINE 1114
 924 -----HLNFNKSVDITVLEQODMMWFGEVQOQKMPFKSVYKLISGP----- 965
 1115 EFPPIQELHIDESSSDDDE-----FEDTRE-----ISATVKTLQTYVNAQPTSSLEIH 1166
 966 ---VAKSISIDTGPFPSPASLKRVASPAKPAIP--GEBFLAMYYESSEGDLTPQOQD 1020
 1167 TEQVIKYPAFGTSPSHNNGNSKKAJSTNSILPVKDEFDDEFAGL-----EQAAVEEDNG-- 1219
 1021 VIVVTKQGDWMTGTVGDKSGVFPSPNYVRLKDSF-----GSGTACKGSLQKREI 1071
 1220 -----ADSESEFENVANAGSEQFETIDHKLDELQNNATGTLTSSSNFTI--PKQV 1272
 1072 AQVIVATATGPEQGLTAPGQILIRKKNPGGMMGELQARQKQKQIPFANYVKLLSP 1131

1273 QQ-----QSTSDPAQV-----SNDEWD-ELFA-----GF----- 1295
 1132 GTSKITPEL--PKTAVQPAVCQVIGMTDYTAQDD--ELAFSGQIIN 1176
 1296 GNSKAEPTKATPSPPOQ-----IPLKNQPIVDASLSKGPLVN 1334

RESULT 10
 528589
 trichohyalin - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
 C:Accession: 528589
 R:Fieltz, M.J.; Rogers, G.E.
 submitted to the EMBL Data Library, December 1992
 A:Description: Examination of the gene encoding rabbit trichohyalin.
 A:Reference number: 528589
 A:Accession: 528589
 A:Molecule type: DNA
 A:Residues: 1-1407 <FIE>
 A:Cross-references: EMBL:Z19092; NID:G1746; PID:CAA79519.1; PID:G1747
 C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she
 Covalent modifications to this protein include conversion of arginine to citrulline and t
 A:introns: 46/3
 C:Superfamily: trichohyalin; calmodulin repeat homology
 C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
 F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 6.2%; Score 391.5; DB 1; Length 1407;
 Best Local Similarity 29.7%; Pred. No. 1.2e-08;
 Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

326 DQRLPESSSEDEQPEKLPVTEDEKKRENFERSVELEKRRQALLQEQ----- 375
 198 EEPTEBQLRRRQEQELKRELEBEOQRREERHERALQEEBQLRQRWRREPREBQ 257
 376 -----RKEORLQLEBAEOKERERQEQEAKROLELEKQRELEKOREBER 426
 258 QLRLEBIEIRERQLEBEBERQQLREBQLEBEBERQQLREBEBERQLEBEBER 317
 427 RKE-----IERE--AAKREL-----ERQOLEMEHNRROELNQNKQEGTVLVKAR 473
 318 RREQLREBEBERQQLREBEBERQLEBEBERQLEBEBERQLEBEBERQLEBEBER 376
 474 RKTLEFELALNDKQHLEGKLDIRCLATQROEISTNKSRELIAETIHLQOQLO-- 531
 377 QROLESEKGAQSKVYSRPRROEQLRQDERQO-----ROERRELEBEOARRQOQWAE 432
 532 -ESQOMGLR-----IPEKQILSDOLKQVOQ-----NSLHDSLLTKRALEAKELARQ 580
 433 EESERRRQLARSLSLEBQRLAEERQEQERFEEBQERERQEOLEFLBEBQLORE 492
 581 LREQLDE-----VERETRSKQLEI--DVFNQOKE--LREIHS-----KQOLQKRS 623
 493 RAQQLQEEBDFQERERRRRRQEQEQRPQGTWQOQEBAQRHRLTYAKPGQEQOLREB 552
 624 LEAARLKQEKERKSLERK--QKEDAKRQVQEDKQK--LEHVOEQOPRRKPHEDR 679
 553 LQREKRRQ--EREREYREBEKQLQREDEKRRQEREROYRELEBRLQEBQDLRLKREBQ 611
 680 L---KREDSVAKKEABERAKPEWQ 700
 612 LQREBERRLRQERERKLRBEEQ 635

RESULT 11
 A45592
 liver seage antigen USA-1 - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jun-2000
 C:Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
 R:Zhu, J.; Hollingdale, M.

submitted to the EMBL Data Library, November 1990

A:Reference number: S24597

A:Accession: S24597

A:Molecule type: DNA

A:Residues: 1-1909 <ZHU>

A:Cross-references: EMBL:X56203; NID:g9915; PID:g9916

R:Zhu, J.; Hollingdale, M.R.

Mol. Biochem. Parasitol. 48, 223-226, 1991

A:Title: Structure of Plasmodium falciparum liver stage antigen-1.

A:Reference number: A45592; MUID:92107224; PMID:1840628

A:Accession: A45592

A:Molecule type: DNA

A:Residues: 1-195;638-688;1165-1215;1590-1909 <ZH2>

A:Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC

R:Guerin-Marchand, C.; Drullhe, P.; Galey, B.; Londono, A.; Pataraotikul, J.; Beaudoin,

Nature 329, 164-167, 1987

A:Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c

A:Reference number: S29393; MUID:87315391; PMID:3306406

A:Accession: S29393

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 323-387 <GUE1>

A:Cross-references: EMBL:M28266

R:Guerin-Marchand, C.; Drullhe, P.; Galey, B.; Londono, A.; Pataraotikul, J.; Beaudoin,

submitted to the EMBL Data Library, April 1992

A:Description: a liver-stage-specific antigen of plasmodium falciparum characterized by

A:Reference number: S34842

A:Accession: S34842

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 323-381, 'HKAL' <GUE2>

A:Cross-references: EMBL:M28266

A:Note: difference at carboxyl end due to frameshift error

C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.

C:Keywords: trichohyalin; calmodulin repeat homology

C:Keywords: EF hand

F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-S-D-L-E-Q-E-R-R)

Query Match 6.1%; Score 383.5; DB 2; Length 1909;
Best Local Similarity 28.3%; Pred. No. 3.6e-08;
Matches 164; Conservative 113; Mismatches 198; Indels 105; Gaps 27;

QY 197 KLOKQAFDVASAPPAEWAPOSRLKRYQLFNSHDKTMSGHLNCPQARTTIMOSLPQ 256
DB 1041 KLOQBOOS-DLEERLAKKLEQEQSDLEERLAKK-----KLOQ-----QOSDLQ 1085
QY 257 AQLAS---IWLSDIDOD-----GKLTAEFFILAM-HLIDVAMSGPLPVPPEYIPPSF 308
DB 1086 ERLAKKLEQEQSDLEERLAKKLEQEQSDLEERLAKKLEQEQ-----QSDL 1134
QY 309 RRVREGSGMSVSISSSSVDQ-----RLPEBPS-EDEQPEKKLPYTFEDKRENPERGS 361
DB 1135 EERLAKKLEQEQSDLEERLAKKLEQEQSDLEERLAKKLEQEQSDLER--TKAS 1191
QY 362 VLEKRRQALTEQORKEQERL-AQLERAEQERKER-EOEBA---KROLELEKOLEKOR 416
DB 1132 KETLOEQSDLEERLAKKLEQEQSDLEERLAKKLEQEQSDLEERLAKKLEQEQS 1251
QY 417 ELERORE-----EERKEIERREBAKREL-EROROLEWERNRROELINORNL-EOEGTV 469
DB 1252 DLEERRAKKLEQEQSDLEERLAKKLEQEQSDLEERLAKKLEQEQSDLEERLAK 1311
QY 470 LKARKKLEFELEALNDKQHGLEGKQDTRCLATQROEISTNSRELRIKEITHTLOOQ 529
DB 1312 EKLQEQSDLEERLAKK-----LQEQSDLEERRAKKLEQEQSDLEER 1358
QY 530 -----LQEQSDLEERRAKKLEQEQSDLEERRAKKLEQEQSDLEER 583
DB 1359 KLAKKLEQEQSDLEERLAKKLEQEQSDLEERRAKKLEQEQSDLEERRAKKLEQ 1418
QY 584 QLEVERETRS--KLQETIDVFNQKEL-REIHSKQOLQORS--LEAARKQKEQERS 638
DB 1419 QOSDLEERRAKKLEQEQ-----QOSDLEERRAKKLEQEQSDLEERRAKKLEQEQS 1472

QY 639 -LE---LEKOKDAQRVQERDKQWLEHVOE-----EQPRPKPH-----EEDRL-- 680
DB 1473 DLEERLAKKLEQEQSDLEERLAKKLEQEQSDLEERRAKKLEQEQSDLEERLAN 1532
QY 681 -KREDVAKKEAEERAKPEMODKQRLPHQPAKLATQ 719
DB 1533 EKLQEQSDLEERRAKKLEQEQSDLEERRAKKLEQEQ 1572

RESULT 12

A:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999

C:Accession: A45973

R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.

J. Biol. Chem. 268, 12164-12176, 1993

A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E

ed (cross-linking) protein.

A:Reference number: A45973; MUID:93280194; PMID:7685034

A:Accession: A45973

A:Molecule type: DNA

A:Residues: 1-1898 <LEE>

A:Cross-references: GB:L09190; NID:g292835; PID:AAA6582.1; PID:g292836

A:Note: authors translated the codon AGG for residue 1714 as Pro

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root st

covalent modifications to this protein include conversion of arginine to citrulline and

C:Keywords:

C:Keywords: GDB:THH

A:Map position: 1921-1921

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 6.1%; Score 381; DB 1; Length 1898;
Best Local Similarity 27.0%; Pred. No. 4.5e-08;
Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

QY 326 DQRLPEBPSDEEQPEKKLPYTFEDKRENPERGSVLEKRRQALBQ--RKEQ--R 381
DB 348 EQQLRE---QEERREQQLRREQEERREQQLRREQEERREQQLRREQQLR 404
QY 382 LAQLERAEQERKER-ROEQAKR--QLELEKOLEKOLERQER-- 426
DB 405 EQQLREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 464
QY 427 ---RKEIRREAKKELERQOLLEWRNRQOLNQRKKEQEGTVVLAKRKLEFELEA 483
DB 465 RLKREQERRDWLKREESTERN--EQERRQOL-KRQEEB-----RREKWLKEEB- 513
QY 484 LNDKQHGLEGKQDTRCLATQROEISTNSRELRIKEITHTLOOLESQ-----QWL- 537
DB 514 ---ERREQERREQQLR-REGEERRRQRLKQEEERLQQLRSEQQLRREQEERLEQLK 570
QY 538 -----GLIPEKQILSPQ---KQVQNSLHRDSLTTLKRALFAKLARQ 580
DB 571 REERKLEQERREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 630
QY 581 LREQLD--VERETRSKLQETIDVFNQKELR---EHSKQOLQORS---EAARKQ 632
DB 631 QERREDEKRLKREPEERREHLLSEQERREHQLRREQQLRREQQLRREQQLR 690
QY 633 -----EQERKSLLEL-EKQEDAQERVQERDKQWLEHVOEOPRPRK----- 673
DB 691 LKREHEERREDELAEEQEQARERIKSRIPWQQLSEADARQSVLLEAPQAGABA 750
QY 674 PHEERLRLKREDSVRRKKEAEERAKPEMODKQSLRPHQPAKLATQAWSTTEKPLTIS 733
DB 751 POEQEKRRRESELOMOEERRAHROQEQEERDFTWQAE-----EKSEGRORIS 803
QY 734 AQESVAVVYVYALYPFESRSHDE 756

Db 804 ARPLRERERQOLAEERQREQ 826

RESULT 13

A40691
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Sep-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: A40691; A34209; S32633
R:Field: M.J.; McLaughlin, C.J.; Campbell, M.T.; Rogers, G.E.
J:Cell Biol. 121, 855-865, 1993
A:Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding
A:Reference number: A40691; MUID:93260018; PMID:7684041
A:Accession: A40691
A:Molecule type: DNA
A:Residues: 1-1549 <F1E>
A:Cross-references: EMBL:Z18361; NID:G295940; PIDD:CA79165.1; PID:G295941
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covariant modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:introns: 46/3
A:Note: single copy gene
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>
F:387-851/Region: 28-residue repeats
F:886-1519/Region: 23-residue repeats

Query Match 6.0%; Score 375.5; DB 1; Length 1549;
Best Local Similarity 27.6%; Pred. No. 5.9e-08;
Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;
332 EPSSSEDEQPEKPLPVTFE-----DKRENFGSVLEKRRQALLQOEKQERLAQER 387
Db 383 KPAGREVRREBQRLKKEKIQREKRRERQYREVLRNEF--ERLQREBQLOREER 440
Qy 388 AEQRKERERQ-----EQAKROLELEKQ-----LEKQRELERQREER 426
Db 441 EKRRERERQYLEKVELMEBEOLOREBERERQREKQYLEKVELREBEOLOREERKR 500
Qy 427 RKELEERRAAGRELERQOLEMERNRQELLNQRNKEBEGTVLKARKTLEFELALND 486
Db 501 ROEBERQYLEKVELQOEBOLOREERK-----RQRE-----RQYLE----- 538
Qy 487 KQHOLEGLQDIRCLATROEISTNKSRELIAETIHLQ--OOLQESQMLGRLLPEK 544
Db 539 -KVELQEEQ-----LQREKERERQREKQYLEKVELQOEBOLOREKQREER 590
Qy 545 QILSDQLKQVOQNSLHRDLSLLTKRALAKELARQOLEQULDEVERTSKLOEDIVFNN 604
Db 591 QYL-EKVELQEEBOLOREERKQREERQYLEKVELQOE--EYVORQREK-----RR 641
Qy 605 QLKRLREHSQOQOKQKSLAALUKQOEKRSLELEK-----OKEDAQR 650
Db 642 QERROYLEKLEQOEBOLOREERQLEREERKQREERQYLEKVELQOEBOLOREERK 701
Qy 651 RVQERDQKML--ENVQEOEPRPKPHEDRLKREDSVRKKAER--AKPEMODKQRL 706
Db 702 RROEREROYLEKELOQOE--RLQREKEQLOQREDRKQVAKQVARKTLEBELQOEERL 758

RESULT 14
T15597

hypothetical protein C25A11.4b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15597
R:Favell, T.
submitted to the EMBL Data Library, October 1995
A:Description: The sequence of C. elegans cosmid C25A11.
A:Reference number: Z18375
A:Accession: T15597
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-737 <F8AV>
A:Cross-references: EMBL:U39650; NID:G1049376; PID:G1049381; PIDD:AAA80392.1; CESP:C25A11
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:C25A11.4b
A:introns: 65/3; 88/3; 126/3; 433/2; 494/2; 711/2
Query Match 5.7%; Score 359.5; DB 2; Length 737;
Best Local Similarity 23.6%; Pred. No. 1.1e-07;
Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;
Qy 181 KSSSFSRSGSQSLNTKLOQAQFVVASAPPAEMAVPOSSRLKYRQFNSS----HDKT 235
Db 7 QSISSRHEDLSBHATSRSTVRIPVHAPSTAP---SHSSVFEYHMMPTTSTYHHVET 63
Qy 236 MSG--HITGQATILMQS--SLPQALASINWLSLDIDQKLTAEFILAMHLIDVAMS 291
Db 64 PSEBYRREVTITIRSTALSTALSTPL----- 91
Qy 292 GQPLPVLPEYIP-PSFRVRSGSGSVISSSVDQRL-----PEPSSEDE---Q 339
Db 92 GRPASPL--DRYLPYPTTTTSGDGT--REKTVDYKYVHNDIEQERIKEDQARRQ 148
Qy 340 QPEKPLPVTEFDKRENFGSVLEKRRQ-----ALLQQRKQERLAQERAEQER 392
Db 149 QEBQDRDRDNRRIQAQREHQMEREERQQLSERALERRADRERLQ--ERLRRQ 207
Qy 393 KEREERQ-----EAKROLELEKQOLELERQREERERKEIEREAKRELER 442
Db 208 REKKRRERWRLSIRLAEBEALARRALEKER-IDREVAEERKTEWERERARLER 266
Qy 443 QROLERNRNQ-----ELNQNKQEOGTV---VLKARKTLEFELALNDKKGLEGK 495
Db 267 ER-LIEERROKQKAEIRIRERREHRIETIRIKRIRIRERERREKKAEBD--- 322
Qy 496 QDIRCLATROEISTNKS--RELIAETIHLQOOLQESQMLGRLLPEKQILSDQLKQV 554
Db 323 ---RLRRLRLERIERERRELERERQELQREAEADR-----ERQRLDEARER 371
Qy 555 -QQNSLHRDLSLLTKRALAKELARQOLEQULDEVERTSKLOEDIVFNNQ--KELEI 612
Db 372 RREERERRELVADVHRQAEERERLKQREBEAELEIRILEQOKIDMERIDAERER 431
Qy 613 HSKQOLQKSL--EAALUKQOEKRSL---ELEKQEDQKQVQERDQKMLSHVQOEQ 668
Db 432 ERKEERERRELEIRAKRKKAERDRDLDEMERVAREERERKEKREBERRIIAEKER 491
Qy 669 PRPKPHEE-----DLKREDSVRKKAERAKPEMODKQSL 706
Db 492 KRQOEEREEIARINELQRAAARQANALDRQOQDELDRLK-AQELSEREMREKERRD 550
Qy 707 FHHQEPKATATQAPMSTTEKGPLTISAQSVYVYRALYPPESSHDTITQPGDIVM 766
Db 551 REBANEAQALADLLERERHNLQIRENERRAVERANNRLE--DRSRDLT-----DIV 603
Qy 767 VDSQTEPEPWLGGELKGTGWFPANYAEKIPNEVEYTPAKPTDILTSAPAKLARETP 826
Db 604 RERSSEKQ-----FELEKRLIAKEAMNRKK--NHLSSS---TLAKLT 643
Qy 827 APPLVTSSEPTTP-----NNWADFSSTWSSS 854
Db 644 QPMYVTTREDEVTTKVERQVIERIDRWVWVEDVPAVSQS 683

RESULT 15

T15598
hypothetical protein C25A11.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15598

R: Favello, T.
submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid C25A11.

A:Reference number: Z18375

A:Accession: T15598

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1017 <FAV>

A:Cross-References: EMBL:U39650, NID:G1049376, PID:G1049380, PIDN:AAA80391.1; CESP:C25A11

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C25A11.4a

A:Introns: 65/3; 86/3; 126/3; 433/2; 494/2; 711/2; 735/1; 792/3; 833/3; 873/2

Query Match 5.7%; Score 359.5; DB 2; Length 1017;

Best Local Similarity 23.6%; Pred. No. 1.6e-07;

Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;

QY 181 KSSFSRSGSGSOLNKLKQKQSPDVASAPPAEMAVPQSSRLKYLQFNSE-----HDKT 235

DB 7 QSISSREDLSEHATSSTVREIPVHRAPSTAP--SHSVFEYHMPPTTSTYHVT 63

QY 236 MSG--HLTGPQARTILMQS--SLPQQLASIWNLSDIDQDKLTAEFILAMHLIDVAMS 291

DB 64 PSDEYRRREVMTITITRSTALSCPTL----- 91

QY 292 GQPLPPVLPPEYIP-PSFRRVSGSGSVISSSVDQRL-----PEPSSDE--Q 339

DB 92 GRPASPL--DRYLPYPTTTTSGDRT-REKTVVYKTYHRDIEQERRIREDOARQ 148

QY 340 QPEKLPVTFEDKKRNFGRGVELEKRO-----ALLEQKQERLAQLEPAQER 392

DB 149 QEEQDRDRDNARILAQREHMERLEEQNLSEALAEERRADKERLQ-ERLLRQ 207

QY 393 KERERQEQ-----EAKRLLEKQLEKQERLEKQERREERREKREAKRELER 442

DB 208 REKKRREEMDRLESTRLAEAEALARRALEKER-IDREKAEERKTMERLERERARLER 266

QY 443 QROLEWERNRQ---ELLQNRKQEGTV---VLKARRKTLFELEALNDKKGQLEGKL 495

DB 267 ER-LEERERQKKAETERIERERERHERIEIRIKERIEREREREREKKAED--- 322

QY 496 QDIRCLATQROEISTNKS-RELRIAEITHLQOQLOESQOMLGRILPEKQILSDQKOV 554

DB 323 ----RLNERLELEIERERRELEAREQEBELQREAEADR-----ERQLDEARER 371

QY 555 -QONSILHDSLLTLKRALFAKELARQOLREOLDEVERETRSKLOEIDVFNOL-KELREI 612

DB 372 RRRREERRAEVLADVHRQALERERLRKQEREAERLERIRLBOQKIDMERIDERRER 431

QY 613 HSKQOLQKQKSL-EAARLKQKQERKSL--ELEKQEDAQRRVQERDKOMLEHVQOEEO 668

DB 432 ERKEERERREFELIEAARRRKEAARDRLDEMERERVREERERREKRRQERRIAEKER 491

QY 669 PRPKRHEE-----DLKREDSVRKKEAEERAKPEMOKQKSL 706

DB 492 KRQEEEEEIARLNLQRAAAARQAQNAELDRQRODELDRK-AQELSERERREKERD 550

QY 707 FHHPQPAKLATQAPWSTKEGSLTISAQESVYVYRALYPFESRSHDEITQPDIVM 766

DB 551 RERANFEAQLADLLERENQLIRERREAEVBRANKRL-DRRSRDKL-----DHIV 603

QY 767 VDESQTGEPCWLGGELKGTGWFPANYAEKIPENEVPTPAKPYTDLTSAAPAKLARETP 826

DB 604 RERSEKEQ-----FELEKRLAEKEMNRKK--NHLLSSE---TLAKLT 643

QY 827 APLVTSSEPSSTP-----NNWADSSSTWPS 854

DB 644 QPMYTTREPEVTTKVERQVIERIDRWVWEDVPYAFSQS 683

Search completed: December 4, 2003, 15:20:53
Job time : 31 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 4, 2003, 15:04:48 / Search time 15 Seconds

(without alignments)
3802.893 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPTFGSGIDVMAITVE.....VGLFSPSNVYKLTITMDPSQ 1213

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6269	100.0	1714	1 ITN1_MOUSE	Q920R4 mus musculus
2	6021	96.0	1217	1 ITN1_RAT	Q9W99 ratu
3	5842.5	93.2	1721	1 ITN1_HUMAN	Q15811 homo sapien
4	5063.5	80.8	1270	1 ITN1_XENLA	Q42287 xenopus lae
5	3068	48.9	1596	1 ITN2_HUMAN	Q92406 mus sapien
6	3049.5	48.6	1558	1 ITN2_MOUSE	Q92006 mus musculus
7	486.5	7.8	896	1 EP15_HUMAN	P42566 homo sapien
8	486	7.8	897	1 EP15_MOUSE	P42567 mus musculus
9	410.5	6.5	1480	1 PANT_YEAST	P32521 mus musculus
10	403	6.4	1381	1 YBET_YEAST	P34216 saccharomyc
11	391.5	6.2	1407	1 TRHY_RABIT	P37709 oryctolagus
12	381	6.1	1898	1 TRHY_HUMAN	Q07283 homo sapien
13	375.5	6.0	1549	1 TRHY_SHEEP	P22793 ovis aries
14	330	5.3	793	1 CALD_HUMAN	Q05682 homo sapien
15	330	5.3	1239	1 M4K4_HUMAN	Q95819 homo sapien
16	327	5.2	1794	1 YDC9_SCHPO	Q10172 schizosacch
17	323.5	5.2	709	1 SHP1_RAT	Q92569 ratu
18	322.5	5.1	709	1 SHP1_MOUSE	Q81550 mus musculus
19	321	5.0	665	1 SHP1_HUMAN	Q96997 homo sapien
20	314.5	5.0	887	1 YLX8_CAEEL	P46504 caenorhabdi
21	309.5	4.9	2611	1 BPL1_MOUSE	Q91208 mus musculus
22	309	4.9	637	1 C2AP_MOUSE	Q91207 mus musculus
23	306.5	4.9	771	1 CALD_CHICK	P12957 gallus gall
24	304.5	4.9	585	1 INVO_HUMAN	P07476 homo sapien
25	304.5	4.9	958	1 IF3A_TBOAC	Q40554 nicotiana t
26	302	4.8	4684	1 PLE1_HUMAN	Q15149 homo sapien
27	299	4.8	1233	1 M4K4_MOUSE	P97820 mus musculus
28	298	4.8	4473	1 PLE1_CRIGR	Q91155 cricetus
29	296	4.7	4687	1 PLE1_RAT	P30427 ratu
30	294.5	4.7	2442	1 CEP1_HUMAN	Q9B7K3 homo sapien
31	287	4.6	639	1 C2AP_HUMAN	Q9YK6 homo sapien
32	283	4.5	733	1 VINE_MOUSE	Q91228 mus musculus
33	283	4.5	1805	1 NEST_RAT	P21263 ratu

34	281.5	4.5	544	1 INVO_AOTTR	P24708 actus tri
35	280	4.5	522	1 INVO_HYLA	P17941 hylobates 1
36	280	4.5	1338	1 MYS_AEGIR	P24733 aegipacten
37	279	4.5	1023	1 GLT_DROME	P33438 drosophila
38	279	4.5	2779	1 LVA_DROME	Q8651 drosophila
39	278.5	4.4	678	1 GARP_PLAFA	P13816 plasmodium
40	278	4.4	1360	1 CING_XENLA	Q99473 xenopus lae
41	277	4.4	377	1 NCK1_HUMAN	P16333 homo sapien
42	277	4.4	1208	1 DC11_DROME	P18169 drosophila
43	276.5	4.4	1197	1 CING_HUMAN	Q92277 homo sapien
44	276.5	4.4	1300	1 DYNB_NEUCR	Q01397 neurospora
45	276	4.4	407	1 M21_STRPY	P50468 streptococc

ALIGNMENTS

RESULT 1	ID	ITN1_MOUSE	STANDARD	PRT	1714 AA.
AC	Q920R4	Q9R143			
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	15-SEP-2003	(Rel. 42, Last annotation update)			
DE	Intersectin 1 (EH and SH3 domains protein 1).				
GN	ITSN1 OR ITSN OR ESEL.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
NC	NCBI_Taxid=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A. (ISOFORM 1).				
RX	MEDLINE=99164083; PubMed=10064583;				
RA	Senger A.S., Wang W., Bishay J., Cohen S., Egan S.E.;				
RT	"The EH and SH3 domain Esh proteins regulate endocytosis by linking to				
RL	dynamitin and Eps15."				
RL	EMBO J. 18:1159-1171(1999).				
RN	[2]				
RP	SEQUENCE OF 966-1714 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 545-599				
RP	FROM N.A. (ISOFORMS 1 AND 2).				
RC	STRAIN=129/Ola; TISSUE=Spleen;				
RA	Skripkina I.V., Teysa L.O., Anoprienko O.V., Slavov D., Tassone F.,				
RA	Ryndtich A.V., Gardiner K.;				
RL	"Mouse homologues of human chromosome 21 genes."				
RT	Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.				
CC	- FUNCTION: Adapter protein that may provide indirect link between				
CC	the endocytic membrane traffic and the actin assembly machinery.				
CC	May regulate the formation of clathrin-coated vesicles.				
CC	- SUBUNIT: Interacts with dynamitin, SNAP-25 and SNAP-23. Clusters				
CC	several dynamitin in a manner that is regulated by alternative				
CC	splicing. Also binds clathrin-associated proteins and other				
CC	components of the endocytic machinery, such as N-WASP, Eps15 and				
CC	Stonin 2 (By similarity).				
CC	- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.				
CC	- ENRICHED IN synaptoosomes (By similarity).				
CC	- ALTERNATIVE PRODUCTS:				
CC	Event=Alternative splicing; Named isoforms=2;				
CC	Comment=Additional isoforms seem to exist;				
CC	Name=1; Synonyms=Essell;				
CC	IsoId=Q920R4-1; Sequence=Displayed;				
CC	Name=2;				
CC	IsoId=Q920R4-2; Sequence=VSP 004296;				
CC	- TISSUE SPECIFICITY: Widely expressed. Expressed at high levels in				
CC	brain, heart and skeletal muscle.				
CC	- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,				
CC	bind to dynamitin (By similarity).				
CC	- DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23 (By				
CC	similarity).				
CC	- MISCELLANEOUS: Overexpression results in the inhibition of the				
CC	transferrin uptake and the blockage of the clathrin-mediated				
CC	endocytosis.				
CC	- SIMILARITY: Contains 1 C2 domain.				
CC	- SIMILARITY: Contains 1 DBL-homology (DH) domain.				

CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF132481; AAD19749.1; -
 DR EMBL; AF132478; AAD19746.1; -
 DR EMBL; AF169621; AAD48848.1; -
 DR EMBL; AF356517; AAK40228.1; -
 DR HSSP; P29354; IGFC.
 DR MGD; MGI:1338069; Ican.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR00261; EPS15_homology.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR00219; RhGEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhGEF; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00238; C2; 1.
 DR SMART; SM00054; EH; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhGEF; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2_DOMAIN_2; 1.
 DR PROSITE; PS50010; DH_2; 1.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 FT Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 221 310 EH 2.
 FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLRQ).
 FT DOMAIN 332 662 COILED COIL (POTENTIAL).
 FT DOMAIN 738 799 SH3 1.
 FT DOMAIN 906 964 SH3 2.
 FT DOMAIN 995 1053 SH3 3.
 FT DOMAIN 1067 1131 SH3 4.
 FT DOMAIN 1148 1207 SH3 5.
 FT DOMAIN 1230 1416 PH.
 FT DOMAIN 1455 1564 PH.
 FT DOMAIN 1576 1672 C2 DOMAIN.
 FT DOMAIN 321 324 POLY-SER.
 FT VARSPLC 1214 1714 Missing (in isoform 2).
 FT SEQUENCE 1714 AA; 194284 MW; 4D7AF298397860A7 CRC64;
 Query Match 100.0%; Score 6269; DB 1; Length 1714;
 Best Local Similarity 100.0%; Pred. No. 3.9e-232;
 Matches 1213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 IVALADNNNDGRMDQVEFSIAMKLIKLGQYQUPSTLPVPMKQOYVAISSAPFGIGCI 120
 DB 61 IVALADNNNDGRMDQVEFSIAMKLIKLGQYQUPSTLPVPMKQOYVAISSAPFGIGCI 120
 QY 121 ASMEPLTAVAVPMGSIPIVVGMSPPVSVVPAAPVPLANGAPPVIOPLPAFAHPATWP 180
 DB 121 ASMEPLTAVAVPMGSIPIVVGMSPPVSVVPAAPVPLANGAPPVIOPLPAFAHPATWP 180
 QY 181 KSSSFSSSGPSQNTLTKQKQSPDVASAPPAAMVAPQSSRLTKYRQLFNSHDTMSGHL 240
 DB 181 KSSSFSSSGPSQNTLTKQKQSPDVASAPPAAMVAPQSSRLTKYRQLFNSHDTMSGHL 240
 QY 241 TGPARITIMQSSLPQAKLASIMVNLSPIDQDKLTAEFFIAMLIDVAMSGQLPVLP 300
 DB 241 TGPARITIMQSSLPQAKLASIMVNLSPIDQDKLTAEFFIAMLIDVAMSGQLPVLP 300
 QY 301 PEYIPPSFRRVRSQSGMSVSISSSVQDRLPEPSSSEDEQPEKKLPVTFEDKKENFERG 360
 DB 301 PEYIPPSFRRVRSQSGMSVSISSSVQDRLPEPSSSEDEQPEKKLPVTFEDKKENFERG 360
 QY 361 SVELKRRQALLBQORKEQERLAQLEBAEQRRKEREQEQAKQLELEKQLEKORELER 420
 DB 361 SVELKRRQALLBQORKEQERLAQLEBAEQRRKEREQEQAKQLELEKQLEKORELER 420
 QY 421 QREBERRKEIRBRPAKRELERORQLEMERRRQELNQRKQEQGTIVLAKARKTLEFE 480
 DB 421 QREBERRKEIRBRPAKRELERORQLEMERRRQELNQRKQEQGTIVLAKARKTLEFE 480
 QY 481 LEALNDKKGQLEGKLODIRCLATQROEISTNKSREIRIAEITHLQOQLQESQOMIGRL 540
 DB 481 LEALNDKKGQLEGKLODIRCLATQROEISTNKSREIRIAEITHLQOQLQESQOMIGRL 540
 QY 541 IPEKQILSDQKQVQNSLHRDLSLTLLKRALEAELARQUREQDVERERTSKLOEID 600
 DB 541 IPEKQILSDQKQVQNSLHRDLSLTLLKRALEAELARQUREQDVERERTSKLOEID 600
 QY 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEQERSLELEKQEDAQRRQERDQWL 660
 DB 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEQERSLELEKQEDAQRRQERDQWL 660
 QY 661 EHVQOEBQPRPKHEEDRLKREDSVKKKEAEBAKEMODKQRLPHPHQEPKATQA 720
 DB 661 EHVQOEBQPRPKHEEDRLKREDSVKKKEAEBAKEMODKQRLPHPHQEPKATQA 720
 QY 721 PWSITTEKGLTISQOESVKKVYVYALYPFESRSHDEITIQGDIVMDESGTGESPGWLG 780
 DB 721 PWSITTEKGLTISQOESVKKVYVYALYPFESRSHDEITIQGDIVMDESGTGESPGWLG 780
 QY 781 ELKQKTGMPFANYAEKIPENEVPTPAKPVTDLTSAAPKTLARETPAPLPYTSSEPSITP 840
 DB 781 ELKQKTGMPFANYAEKIPENEVPTPAKPVTDLTSAAPKTLARETPAPLPYTSSEPSITP 840
 QY 841 NNMADFSSTWSSSNEKRETDNMDTMAAQPSTLVPSAQQLQORSAFTPATATGSSPSFVL 900
 DB 841 NNMADFSSTWSSSNEKRETDNMDTMAAQPSTLVPSAQQLQORSAFTPATATGSSPSFVL 900
 QY 901 GQGERKVEGLQALYIPRAKKNHNLNFKSVITVLVQOQDMWMMGCVQCGMPKSVK 960
 DB 901 GQGERKVEGLQALYIPRAKKNHNLNFKSVITVLVQOQDMWMMGCVQCGMPKSVK 960
 QY 961 LISGPVRKSTSIDTGPTEPSASLKRVASPAKPAIPGEFTAMTYTESSEGGDTLFOGD 1020
 DB 961 LISGPVRKSTSIDTGPTEPSASLKRVASPAKPAIPGEFTAMTYTESSEGGDTLFOGD 1020
 QY 1021 VIIVTKKDGDMWTGVDKSGVFPNSNYVRLKDSGSGTAGTGSGLGKKPEIAQVIASYAA 1080
 DB 1021 VIIVTKKDGDMWTGVDKSGVFPNSNYVRLKDSGSGTAGTGSGLGKKPEIAQVIASYAA 1080
 QY 1081 TGPEQLTLAPQQLILIRKKNPGMWEGLQARKKROIGWPPANVYKLLSGTSTKITPTE 1140
 DB 1081 TGPEQLTLAPQQLILIRKKNPGMWEGLQARKKROIGWPPANVYKLLSGTSTKITPTE 1140
 QY 1141 LPKTAVQAVCVIGMTDYTAQNDDDELAFSGQIITVNLNKEDPMMWKEVSGVGLFPSN 1200

Db 1141 LEKTAIVQPAVCCVIGDYTAQNDDELAFSKQIIVLTKNEPDMWKEVSGVGLFPSPN 1200
 QY 1201 YKLTITDMPDSCQ 1213
 Db 1201 YKLTITDMPDSCQ 1213

RESULT 2
 ITN1 RAT
 ID ITN1 RAT STANDARD; PRT; 1217 AA.
 AC Q9WVE9; Q9WVE1; (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).
 GN ITSN1 OR ITSN OR EHSN1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain;
 RX MEDLINE=99303609; PubMed=10373452;
 RA Okamoto M., Schuch S., Suedhof T.C.;
 RT "EHSN1/intersectin, a protein that contains EH and SH3 domains and binds to dynamin and SNAP-25. A protein connection between exocytosis and endocytosis?";
 RT J. Biol. Chem. 274:18446-18454(1999).
 CC -1- FUNCTION: Adapter protein that may provide indirect link between the endocytic membrane traffic and the actin assembly machinery. May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters several dynamin in a manner that is regulated by alternative splicing. Also binds clathrin-associated proteins and other components of the endocytic machinery, such as N-Wasp, Eps15 and Stonin 2 (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
 CC -1- ENRICHED IN synaposomes.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q9WVE9-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9WVE9-2; Sequence=VSP_004297;
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
 CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains, bind to dynamin.
 CC -1- DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF127798; AAD30271.1; -;
 DR EMBL, AF134672; AAD31026.1; -;
 DR HSSP, P29354; IGFC.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000261; EPS15_homology.
 DR InterPro: IPR001452; SH3.
 DR Pfam, PF00036; ehnd, 3.
 DR Pfam, PF00018; SH3; 5.
 DR ProDom, PD000066; SH3; 5.
 DR SMART, SM00054; Eph; 2.

DR SMART, SM00027; EH; 2.
 DR SMART, SM00326; SH3; 5.
 DR PROSITE, PS00018; EF_HAND; 2.
 DR PROSITE, PS50031; EH; 2.
 DR PROSITE, PS50002; SH3; 5.
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF_HAND 1 (POTENTIAL).
 FT DOMAIN 221 310 EH 2.
 FT CA BIND 267 279 EF_HAND 2 (POTENTIAL).
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLERQ).
 FT DOMAIN 350 670 COILED COIL (POTENTIAL).
 FT DOMAIN 737 803 SH3 1.
 FT DOMAIN 910 968 SH3 2.
 FT DOMAIN 959 1057 SH3 3.
 FT DOMAIN 1071 1135 SH3 4.
 FT DOMAIN 1152 1211 SH3 5.
 FT DOMAIN 321 324 POLY-SER.
 FT VARSPLIC 1003 1073 Missing (in isoform 2).
 FT /FTID=VSP_004297.
 SQ SEQUENCE 1217 AA; 137154 MW; 6C1323BAE5A5B34B CRC64;

Query Match 96.0%; Score 6021; DB 1; Length 1217;
 Best Local Similarity 95.7%; Pred. No. 7,9e-223;
 Matches 1166; Conservative 19; Mismatches 27; Indels 6; Gaps 2;

QY 1 MAQFTPPGSGSLDVAITVEERAKHDQFLSLKPIAGFTTGDQARNFFQSGLPQVLAQ 60
 Db 1 MAQFTPPGSGSLDVAITVEERAKHDQFQSLKPIAGFTTGDQARNFFQSGLPQVLAQ 60
 QY 61 IVALADNMNDGMDQVERSIAMKLIKLGQGLPSTLPVWKKQPPVAISSAPAGIGGI 120
 Db 61 IVALADNMNDGMDQVERSIAMKLIKLGQGLPSTLPVWKKQPPVAISSAPAGIGGI 120
 QY 121 ASMPPLTVAVAPVPMGSIPIVVGKSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 Db 121 ASMPPLTVAVAPVPMGSIPIVVGKSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 QY 121 AGMPPLTVAVAPVPMGSIPIVVGKSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 Db 121 AGMPPLTVAVAPVPMGSIPIVVGKSPPLVSSVPPAAPPVLANGAPVPIQPLPAFAHPAATWP 180
 QY 181 KSSSFRSGSGSGLNTKLQKQSPVAVASAPPAEWAAPQSSRLKXKRLFNSHDKTWSGHL 240
 Db 181 KSSSFRSGSGSGLNTKLQKQSPVAVASAPPAEWAAPQSSRLKXKRLFNSHDKTWSGHL 240
 QY 181 KSSSFRSGSGSGLNTKLQKQSPVAVASAPPAEWAAPQSSRLKXKRLFNSHDKTWSGHL 240
 Db 181 KSSSFRSGSGSGLNTKLQKQSPVAVASAPPAEWAAPQSSRLKXKRLFNSHDKTWSGHL 240
 QY 241 TGPQARTILMOSLSLPQALASLWNLSIDIDQSKLTAEFFILAMHLIDVAMSGQPLPVLP 300
 Db 241 TGPQARTILMOSLSLPQALASLWNLSIDIDQSKLTAEFFILAMHLIDVAMSGQPLPVLP 300
 QY 301 PEYIPSPFRARVSSGSGMSVSSSVYDRLPEPSSSEDEQPEKLPVTFEDKRENPFEG 360
 Db 301 PEYIPSPFRARVSSGSGMSVSSSVYDRLPEPSSSEDEQPEKLPVTFEDKRENPFEG 360
 QY 301 PEYIPSPFRARVSSGSGMSVSSSVYDRLPEPSSSEDEQPEKLPVTFEDKRENPFEG 360
 Db 301 PEYIPSPFRARVSSGSGMSVSSSVYDRLPEPSSSEDEQPEKLPVTFEDKRENPFEG 360
 QY 361 SVELERKQALLLEQQRKEQERLAQLERAEOERKEREROEAKROLLEKOLEKRELER 420
 Db 361 SVELERKQALLLEQQRKEQERLAQLERAEOERKEREROEAKROLLEKOLEKRELER 420
 QY 361 NLELEKRRQALLLEQQRKEQERLAQLERAEOERKEREROEAKROLLEKOLEKRELER 420
 Db 361 NLELEKRRQALLLEQQRKEQERLAQLERAEOERKEREROEAKROLLEKOLEKRELER 420
 QY 421 QREERREKIERREAKELEROROLEMERNRQELTQYNNDOGIVLAKARRTLLEFE 480
 Db 421 QREERREKIERREAKELEROROLEMERNRQELTQYNNDOGIVLAKARRTLLEFE 480
 QY 481 LEALNDKQHQLEGLKQDRCRLATQROEISTNSRELRIAEITHLQOOLQESQOMLRL 540
 Db 481 LEALNDKQHQLEGLKQDRCRLATQROEISTNSRELRIAEITHLQOOLQESQOMLRL 540
 QY 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALAEKELAQOLREQLDEVERETRSKLQEID 600
 Db 541 IPEKQILSDOLKOVQNSLHRDSTLLTKRALAEKELAQOLREQLDEVERETRSKLQEID 600
 QY 601 VFNQQLKELEIHSQQLQKQSLLEAALNKQEGEKSLLEKQKEDQKRVQEDKQWL 660
 Db 601 VFNQQLKELEIHSQQLQKQSLLEAALNKQEGEKSLLEKQKEDQKRVQEDKQWL 660
 QY 661 EHVQOEQPPRPKPEEDRLKREDSVRKKEAEERAKPEMDQKOSLFPHPQEPKATQTA 720
 Db 661 EHVQOEQPPRPKPEEDRLKREDSVRKKEAEERAKPEMDQKOSLFPHPQEPKATQTA 720

```

Db      661 EHVQOEQRPRKPHEDDKLKRDESVKKKAEBERAKPEVQDKSLFHPQEPAPKA-QA 719
Qy      721 PMSTTEKPLTISAOESVKVYVYALYFPEBSRSHDEITIQGCDIVM-----VDSSQCEP 775
Db      720 PMFTTEKPLTISAOESAKVYVYALYFPEBSRSHDEITIQGCDIVMKGWVDSSQCEP 779
Qy      776 GMLGELGKGTGMPFANYAEKI PENEVPTPAKPYDLTSAAPKLAETPAIPVTSSE 835
Db      780 GMLGEPGKGTGMPFANYAEKI PENEITPAKPYDLTSAAPKLAETPAIPVTSSE 839
Qy      836 PSTTPNNMADPSSSWSSSNEKETDNDVTAADPSLTVPAGQILRQSAFTPATGSS 895
Db      840 PSTTPNNMADPSSSWSSSNEKETDNDVTAADPSLTVPAGQILRQSAFTPATGSS 899
Qy      896 PSPVIGQGEKVEGLOALYPMRAKKNHLPNNSDVTVEEQDMWFGVQGGKGF 955
Db      900 PSPVIGQGEKVEGLOALYPMRAKKNHLPNNSDVTVEEQDMWFGVQGGKGF 959
Qy      956 KSYVVLISGPVRKSTIDTGPTEPSASLKRVAAPAKPAIPGEFIATYTESSEQDGLT 1015
Db      960 KSYVVLISGPVRKSTIDTGPTEPSASLKRVAAPAKPAIPGEFIATYTESSEHGLT 1019
Qy      1016 FQGGDVIVVTKKDGDMWTGTVGDKSGVFPSSNYVRLKDSGSGTAGKTSLGAKPEIAQVI 1075
Db      1020 FQGGHVIVVTKKDGDMWTGTVGDKSGVFPSSNYVRLKDSGSGTAGKTSLGAKPEIAQVI 1079
Qy      1076 ASYATGEGEQLTLAAGQILIRKKNPGGMEGELQARKKQIQGFPANVYKLSRGTSK 1135
Db      1080 ASYATGEGEQLTLAAGQILIRKKNPGGMEGELQARKKQIQGFPANVYKLSRGTSK 1139
Qy      1136 ITPTLPPTAVOPACQVIGMYDYAQNDELAFSKQIINVANKEDPMMWKGESGQVG 1195
Db      1140 ITPTLPPTAVOPACQVIGMYDYAQNDELAFSKQIINVANKEDPMMWKGESGQVG 1199
Qy      1196 LFPSSNYVRLKTDMDPSQ 1213
Db      1200 LFPSSNYVRLKTDMDPSQ 1217

RESULT 3
ID      ITN1_HUMAN
ITN1_HUMAN STANDARD; PRT; 1721 AA.
AC      Q15811; Q95216; Q9UK60; Q9UNK1; Q9UNK2; Q9UQ92;
DT      01-NOV-1997 (Rel. 35, Created)
DT      28-FEB-2003 (Rel. 41, last sequence update)
DT      15-SEP-2003 (Rel. 42, last annotation update)
DE      Intersectin 1 (SH3 domain-containing protein 1A) (SH3P17).
GN      ITSN1 OR ITSN OR SH3D1A.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID:9606;
RN      [1]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC      TISSUE=Fetal brain;
RX      MEDLINE=99017974; PubMed=9799604;
RA      Antonarakis S.E.; Scott H.S.; Chen H.; Schebesta A.; Rossier C.;
RA      Antonarakis S.E.;
RT      "Two isoforms of a human intersectin (ITSN) protein are produced by
RT      brain-specific alternative splicing in a stop codon.";
RL      Genomics 53:369-376(1998).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX      MEDLINE=99415290; PubMed=10482960;
RA      Pucharcos C.; Fuentes J.-J.; Casas C.; de la Luna S.; Alcantara S.;
RA      Arbones M.L.; Soriano E.; Estivill X.; Pritchard M.;
RT      "Alu-sequence cloning of human intersectin (ITSN), a putative
RT      multivalent binding protein expressed in proliferating and
RT      differentiating neurons and overexpressed in Down syndrome.";
RL      Eur. J. Hum. Genet. 7:704-712(1999).
RN      [3]
RP      SEQUENCE OF 963-1721 FROM N.A. (ISOFORM 3).
RC      TISSUE=Brain;

```

```

RA      Teyba L.O.; Kvaasha S.M.; Skripkina I.Y.; Anoprienko O.V.; Slavov D.;
RA      Tassone F.; Ryndlich A.V.; Gardiner K.;
RT      "Mouse homologs of human Chromosome 21 genes";
RL      Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN      [4]
RP      SEQUENCE OF 620-1721 FROM N.A. (ISOFORM 3).
RC      TISSUE=Bone marrow;
RX      MEDLINE=98294438; PubMed=9630982;
RA      Sparks A.B.; Hoffman N.G.; McConnell S.J.; Fowlkes D.M.; Kay B.K.;
RT      "Cloning of ligand targets: systematic isolation of SH3 domain-
RT      containing proteins.";
RL      Nat. Biotechnol. 14:741-744(1996).
RN      [5]
RP      ALTERNATIVE SPLICING.
RC      TISSUE=Brain, and Fetal liver;
RX      MEDLINE=21548828; PubMed=11690630;
RA      Pucharcos C.; Casas C.; Nadal M.; Estivill X.; de la Luna S.;
RT      "The human intersectin genes and their spliced variants are
RT      differentially expressed.";
RL      Biochim. Biophys. Acta 1521:1-11(2001).
RN      [6]
RP      GENE MAPPING.
RX      MEDLINE=96127038; PubMed=9465890;
RA      Chen H.; Antonarakis S.E.;
RT      "The SH3D1A gene maps to human chromosome 21q22.1-->q22.2.";
RL      Cytogenet. Cell Genet. 78:213-215(1997).
CC      -1- FUNCTION: Adapter protein that may provide indirect link between
CC      the endocytic membrane traffic and the actin assembly machinery.
CC      May regulate the formation of clathrin-coated vesicles. Isoform 1
CC      could be involved in brain-specific synaptic vesicle recycling.
CC      -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
CC      several dynamin in a manner that is regulated by alternative
CC      splicing. Also binds clathrin-associated proteins and other
CC      components of the endocytic machinery, such as N-WASP, Eps15 and
CC      Sec23 (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC      Enriched in synaptosomes (By similarity).
CC      -1- ALTERNATIVE PRODUCTS:
CC      Event-Alternative splicing: Named isoforms=4;
CC      Comment=Additional isoforms seem to exist. Alternative splicing
CC      affects domains involved in protein recognition and thus may
CC      play a role in selecting specific interactions;
CC      Name=1; Synonym=Long; ITSN-1;
CC      IsoId=Q15811-1; Sequence=Displayed;
CC      Name=2; Synonym=Short; ITSN-S;
CC      IsoId=Q15811-2; Sequence=VSP_004295;
CC      Name=3; Synonym=Short 2, SH3P17;
CC      IsoId=Q15811-3; Sequence=VSP_004293, VSP_004294, VSP_004295;
CC      Name=4;
CC      IsoId=Q15811-4; Sequence=VSP_004294;
CC      -1- TISSUE SPECIFICITY: Ubiquitous in adult and fetal tissues, except
CC      isoform 1 which is expressed almost exclusively in the brain.
CC      Highly expressed in skeletal muscle, heart, spleen, ovary, testis
CC      and all fetal tissues tested. Expressed at lower levels in thymus,
CC      blood, lung, liver and pancreas. Isoform 1 is expressed in all
CC      brain regions; not expressed in the spinal cord.
CC      -1- DOMAIN: SH3-3, SH3-4 and SH3-5 but not SH3-1 and SH3-2 domains.
CC      bind to dynamin (By similarity).
CC      -1- DOMAIN: The KLRQ domain binds to SNAP-25 and SNAP-23 (By
CC      similarity).
CC      -1- DISEASE: Overexpressed in brain from Down syndrome fetuses
CC      suggesting a gene dosage-dependent contribution to the
CC      abnormalities of Down syndrome.
CC      -1- SIMILARITY: Contains 1 C2 domain.
CC      -1- SIMILARITY: Contains 1 DBP-homology (DH) domain.
CC      -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC      -1- SIMILARITY: Contains 1 PH domain.
CC      -1- SIMILARITY: Contains 5 SH3 domains.
CC      -1- SIMILARITY: Contains 2 EH domains.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL AF064244; AAC78611.1; -;
DR EMBL AF064247; AAC80437.1; -;
DR EMBL AF064245; AAC80437.1; JOINED.
DR EMBL AF064246; AAC80437.1; JOINED.
DR EMBL AF064243; AAC78610.1; -;
DR EMBL AF114488; AAD29953.1; -;
DR EMBL AF114487; AAD29952.1; -;
DR EMBL AF180522; AAD53183.1; -;
DR EMBL U61166; AAC50592.1; ALT_INIT.
DR PDB: 1K11; 29-MAY-02.
DR Genew: HGNC:6183; ITSN1.
DR MIM: 602442; -;
DR GO: GO:0005509; F:calcium ion binding activity; NAS.
DR GO: GO:0005085; F:guanylyl-nucleotide exchange factor activity; NAS.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; Eps15_homology.
DR InterPro: IPR001018; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR00219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF000168; C2; 1.
DR Pfam: PF00036; eHand; 3.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00018; SH3; 5.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 5.
DR SMART: SM00239; C2; 1.
DR SMART: SM00054; EFh; 2.
DR SMART: SM00027; EH; 2.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR SMART: SM00326; SH3; 5.
DR PROSITE: PS50004; C2_DOMAIN_2; 1.
DR PROSITE: PS50010; DH_2; 1.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS50031; EH; 2.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50002; SH3; 5.
DR EndoCytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
KW Alternative splicing; 3D-structure.
FT DOMAIN 21 109 EH 1.
FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
FT DOMAIN 221 310 EH 2.
FT CA BIND 267 278 EF-HAND 2 (POTENTIAL).
FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLER).
FT DOMAIN 351 705 COILED COIL (POTENTIAL).
FT DOMAIN 740 806 SH3 1.
FT DOMAIN 913 971 SH3 2.
FT DOMAIN 1002 1060 SH3 3.
FT DOMAIN 1074 1138 SH3 4.
FT DOMAIN 1155 1214 SH3 5.
FT DOMAIN 1237 1423 DH.
FT DOMAIN 1462 1571 PH.
FT DOMAIN 1583 1679 C2 DOMAIN.
FT VARSPLIC 770 774 Missing (in isoform 3).
FT VARSPLIC 1006 1076 /FTId=VSP_004293.
FT VARSPLIC 1221 1721 Missing (in isoform 3 and isoform 4).
FT VARSPLIC 1221 1721 /FTId=VSP_004294.
FT CONFLICT 114 114 /FTId=VSP_004295.
FT CONFLICT 1088 1088 P -> A (IN REF. 2).
FT CONFLICT 1109 1109 T -> A (IN REF. 3).
FT CONFLICT 1137 1137 G -> R (IN REF. 3).
FT CONFLICT 1361 1361 S -> N (IN REF. 1).
FT CONFLICT 1361 1361 E -> A (IN REF. 2).

FT CONFLICT 1474 1474 N -> S (IN REF. 2).
SQ SEQUENCE 1721 AA; 195532 MW; 8189DDA0F145B4B5 CRC64;
Query Match 93.2%; Score 5842.5; DB 1; Length 1721;
Best Local Similarity 92.5%; Pred. No. 7.1e-216;
Matches 1128; Conservative 39; Mismatches 46; Indels 7; Gaps 3;
QY 1 MAQPTPGGSLDVVAITVEERAKHDOFLSKPIAGFTTGOQANFFQSGLPQVLAQ 60
DB 1 MAQPTPGGSLDVVAITVEERAKHDOFLSKPIAGFTTGOQANFFQSGLPQVLAQ 60
QY 61 IVALADMMNDGMDQVERSIAMKLIKLIKQYQLPSTLPVVKQOPVAISSAPAGIGSI 120
DB 61 IVALADMMNDGMDQVERSIAMKLIKLIKQYQLPSTLPVVKQOPVAISSAPAGIGSI 120
QY 121 ASMPPLTAVAPVPMGSIPIVVGMSPTLVSSVPPAAVPLANGAPPIYQPLPAFHPAATLP 180
DB 121 ASMPPLTAVAPVPMGSIPIVVGMSPTLVSSVPPAAVPLANGAPPIYQPLPAFHPAATLP 180
QY 181 KSSFSRSGPSGSLNTKLOKQSPVAPAPAAEAVQSSRLKXROLFNSHDKTMSGL 240
DB 181 KSSFSRSGPSGSLNTKLOKQSPVAPAPAAEAVQSSRLKXROLFNSHDKTMSGL 240
QY 241 TGPQARTILMOSLTPQAOLASITWNLSDIDQDKLTAEFTILAMHLIDVAMSGQPLPVLP 300
DB 241 TGPQARTILMOSLTPQAOLASITWNLSDIDQDKLTAEFTILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPPSRRRVAGSGGMSYISSYDQRLPEPSSSEDEQOP-EKLLPTFEDKKBENFER 359
DB 301 PEYIPPSRRRVAGSGGMSYISSYDQRLPEPSSSEDEQOP-EKLLPTFEDKKBENFER 359
QY 360 GSVELKKRQALKEORKEOERLAQERAEORKEOERKEOERKOLEKLEKORELE 419
DB 360 GSVELKKRQALKEORKEOERLAQERAEORKEOERKEOERKOLEKLEKORELE 419
QY 420 RQREERERKEIERRAARELERQRLQEMERNRROELNORKEOEGTVLAKARKTLEF 479
DB 420 RQREERERKEIERRAARELERQRLQEMERNRROELNORKEOEGTVLAKARKTLEF 479
QY 480 ELEALNDKXKQLEGLDQIRCLATROEIESTNSKREIRLAEITHLQOOLQESQOMIGR 539
DB 480 ELEALNDKXKQLEGLDQIRCLATROEIESTNSKREIRLAEITHLQOOLQESQOMIGR 539
QY 540 LIPEKQIISDQKQVQNSLRHDSLTTLKRALAEALRQQLRDQDEVKTRKSLQEI 599
DB 540 LIPEKQIISDQKQVQNSLRHDSLTTLKRALAEALRQQLRDQDEVKTRKSLQEI 599
QY 600 DVFNQQLKELEIHSHKQOLQKORSLAARLKQKEQERKSLLEKQEDQORVQBRDKQW 659
DB 600 DVFNQQLKELEIHSHKQOLQKORSLAARLKQKEQERKSLLEKQEDQORVQBRDKQW 659
QY 660 LEHVQOE-EQPPRPKPHEDRIKREDSYRKKEAEERAKPENQDKOSRLPHHQEPAKLAT 718
DB 660 LEHVQOE-EQPPRPKPHEDRIKREDSYRKKEAEERAKPENQDKOSRLPHHQEPAKLAT 718
QY 719 QAPWSTKRGPLTISAQSNVAVYVYRALYPPESRSHDITTOGPDIW-----VESQSG 773
DB 719 QAPWSTKRGPLTISAQSNVAVYVYRALYPPESRSHDITTOGPDIW-----VESQSG 773
QY 721 QAPWSTKRGPLTISAQSNVAVYVYRALYPPESRSHDITTOGPDIWVGKGVESQSG 780
DB 721 QAPWSTKRGPLTISAQSNVAVYVYRALYPPESRSHDITTOGPDIWVGKGVESQSG 780
QY 774 EFGWLGELGKGTGFWPANYAEKIPENEVPTPAKVTLTSAAPAKLAREPAPLPYTS 833
DB 774 EFGWLGELGKGTGFWPANYAEKIPENEVPTPAKVTLTSAAPAKLAREPAPLPYTS 833
QY 781 EFGWLGELGKGTGFWPANYAEKIPENEVPTPAKVTLTSAAPAKLAREPAPLPYTS 840
DB 781 EFGWLGELGKGTGFWPANYAEKIPENEVPTPAKVTLTSAAPAKLAREPAPLPYTS 840
QY 834 SEPSTTPNNWADFSSTWPSSSNEKETDNWDTMAAQPSTLPVSAGQLQORSAPATATG 893
DB 834 SEPSTTPNNWADFSSTWPSSSNEKETDNWDTMAAQPSTLPVSAGQLQORSAPATATG 893
QY 841 SEPSTTPNNWADFSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTW 900
DB 841 SEPSTTPNNWADFSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTWPSSTW 900
QY 894 SSPSVLQGEKVEBGLQALYPMWAKKDNHLPKPSKVITVLEQDDMMWREVGQKGM 953
DB 894 SSPSVLQGEKVEBGLQALYPMWAKKDNHLPKPSKVITVLEQDDMMWREVGQKGM 953
QY 901 SSPSVLQGEKVEBGLQALYPMWAKKDNHLPKPSKVITVLEQDDMMWREVGQKGM 960
DB 901 SSPSVLQGEKVEBGLQALYPMWAKKDNHLPKPSKVITVLEQDDMMWREVGQKGM 960
QY 954 FPKSVYKLISGVRKSTSIDGPTESPASLKRVASPAKPAIPGEFTIAMYTESSQGD 1013
DB 954 FPKSVYKLISGVRKSTSIDGPTESPASLKRVASPAKPAIPGEFTIAMYTESSQGD 1013

Db 961 PFKSVYKLSGPIRKSTMSDSSSPASLKRVASPAKPVVSGEEFIAMTYESSECD 1020
 QY 1014 LTFQGGDYVYTKDGDWMTGTGDKSVFSPSNYRLKDSGSGTAGTSLGKKPEIAQ 1073
 Db 1021 LTFQGGDYVYTKDGDWMTGTGDKSVFSPSNYRLKDSGSGTAGTSLGKKPEIAQ 1080
 QY 1074 VIASAAATGPOLTLAPGQLILIRKNGGWMGELQARGKKRQIGWFPAYVYKLSPT 1133
 Db 1081 VIASATATGPOLTLAPGQLILIRKNGGWMGELQARGKKRQIGWFPAYVYKLSPT 1140
 QY 1134 SKITPTLPKTAQVAVPACVIGMDYDTAONDELAFSGKGIIVNLKEDPPMKGEVSGQ 1193
 Db 1141 SKITPTLPKSTALAAVCQVIGMDYDTAONDELAFSGKGIIVNLKEDPPMKGEVSGQ 1200
 QY 1194 VGLFSPSNYKLTTPMDPSQ 1213
 Db 1201 VGLFSPSNYKLTTPMDPSQ 1220

RESULT 4
 ITN1_XENLA STANDARD; PRT; 1270 AA.
 AC 04287;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Intersectin 1.
 GN ITSN1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 CC Xenopodidae; Xenopus.
 CX NCBI_TaxID=8335;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Oocyte;
 RX MEDLINE=99030416; PubMed=9813051;
 RA Yamauchi M., Hoffman N.G., Hardison N.L., McPherson P.S.,
 RT "Intersectin, a novel adaptor protein with two eps15 homology and five
 src homology 3 domains";
 RL J. Biol. Chem. 273:31401-31407(1998).
 CC -1- FUNCTION: Adaptor protein that may provide indirect link between
 the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF032118; AAC73068.1; -;
 DR PIR, T09194; T09194.
 DR HSSP, P29355; 1SEM.
 DR InterPro, IPR002048; EF-hand.
 DR InterPro, IPR000261; Eps15_homology.
 DR InterPro, IPR00108; Neu_cyt_fact_2.
 DR InterPro, IPR001452; SH3.
 DR Pfam, PF00036; ehand; 2.
 DR Pfam, PF00018; SH3; 5.
 DR PRINTS, PR00499; P67PHOX.
 DR PRINTS, PR00452; SH3DOMAIN.
 DR ProDom, PD000066; SH3; 5.
 DR SMART, SM00054; Eph; 2.
 DR SMART, SM00027; EH; 2.

DR SMART, SM00326; SH3; 5.
 DR PROSITE, PS00018; EF_HAND; 2.
 DR PROSITE, PS50031; EH; 2.
 DR PROSITE, PS50002; SH3; 5.
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.
 FT DOMAIN 21
 FT CA BIND 66 77 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 220 309 EH 2.
 FT CA BIND 266 277 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 325 697 Lys/Leu/Glu/Arg/Gln-Rich (KLERQ).
 FT DOMAIN 349 691 COILED COIL (POTENTIAL).
 FT DOMAIN 732 793 SH3 1.
 FT DOMAIN 897 955 SH3 2.
 FT DOMAIN 986 1044 SH3 3.
 FT DOMAIN 1058 1122 SH3 4.
 FT DOMAIN 1139 1198 SH3 5.
 FT SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A6858 CRC64;
 Query Match 80.8%; Score 5063.5; DB 1; Length 1270;
 Best Local Similarity 80.6%; Pred. No. 2.5e-166;
 Matches 980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;

QY 1 MAQPTPFGSLDVAITVEERAKHQDQLSLKPIAGFITGDQARNFPQSGLPQVLAQ 60
 Db 1 MAQPTPFGSLDVAITVEERAKHQDQFGLKPEAGIITGDQARNFPQSGLPQVLAQ 60
 QY 61 IVALADNMGRMOVEFSTAMKLIKLGQYOLPSTLPVPMKQOVALISA--PARGIG 118
 Db 61 IVALADNMGRMOVEFSTAMKLIKLGQYPLPSTLPVPMKQOVALISA--PARGIG 120
 QY 119 GIAMSPPLTAVAPVPMGSIPIVGMSPPLVSVPPAAVPLPLANGAPVYIQLPAPAHAPAT 178
 Db 121 GIVGIPPLAANAPVPMGSIPIVGMSPPLVSV--TVPLPLNGAPVYIQLPAPAH--SAT 177
 QY 179 WPKSSFSRSGPSQLNTKQAKQSPVVASAPPAEVAVPOSSRLKTRQLFNSHDKTMSG 238
 Db 178 LPKSSFSRSGPSQLNTKQAKQSPVVASAPPAEVAVPOSSRLKTRQLFNSHDKTMSG 237
 QY 239 HLTPQARTIMQSLPQAOQASIWNLSDIDQDGLTAEEFTLMLHLIDVAMSGQPLPV 298
 Db 238 NLTPQARTIMQSLPQAOQASIWNLSDIDQDGLTAEEFTLMLHLIDVAMSGQPLPV 297
 QY 299 LPPEYIPSPFRVNSGSGMSYISSSVYDQRLPEEPPSEDEQPEPKLVTFEDKKRENF 358
 Db 298 LPPEYIPSPFRVNSGSGMSYISSSVYDQRLPEEPPSEDEQPEPKLVTFEDKKRENF 357
 QY 359 RGSVELEKRGQALLIQKQEDERLAQLBRAQEKERERQEQAKQOLEKQLEKQEREL 418
 Db 358 RGNLELEKRGQALLIQKQEDERLAQLBRAQEKERERQEQAKQOLEKQLEKQEREL 417
 QY 419 EROEEERKEIEERBAKRELEROROLEMRNRROELNORNEOEGTVVLKAKRTLE 478
 Db 418 EROEEERKEIEERBAKRELEROROLEMRNRROELNORNEOEGTVVLKAKRTLE 477
 QY 479 FELALNDKQGLQGLQDLIRICRIATQROEISTNKSRELIATITLQOOLQOESQOMLG 538
 Db 478 FELALNDKQGLQGLQDLIRICRIATQROEISTNKSRELIATITLQOOLQOESQOMLG 537
 QY 539 RLPEKQILSDQLKQVOONSILHDSLLTLKALFAKELIARQOLRQOLDEVERETRQOE 598
 Db 538 RLPEKQILSDQLKQVOONSILHDSLLTLKALFAKELIARQOLRQOLDEVERETRQOE 597
 QY 599 IDVFNNOQKELREIHSQOOLQOKSLEAARKQOEKESLELEKQEDADARQVPERQK 658
 Db 598 IDVFNNOQKELREIHSQOOLQOKSLEAARKQOEKESLELEKQEDADARQVPERQK 657
 QY 659 WLEHVOOEOPRPKPEHEDLTKREDSYRKKEAEERAKPEQDQKQSLRFLHPOEPAKLTAT 718
 Db 658 WQDRVQOE--RYFQDEEKEKEESVQKEVEK--KPELOEKPKNFHPPEPEKGLKG 713
 QY 719 QAPWSTTEKGLTISAQSVKVVYVYRALYPRESSHDEITITQPGDIVVWDESGTEGPEWL 778
 Db 714 QIPWNTTEKGLTIN--QDVKVYVYVYRALYPEDASHSDEITTEPGDIIVWDESGTEGPEWL 772

```

QY 779 GGEIKGTGMPFANYAEKIPENEPVTPAKVPTDLSAPAPKLAARETPAPL-PTSSSPS 837
DB 773 GGEIKGTGMPFANYAEKIPENEPVTPAKVPTDLSAPAPKLAARETPAPL-PTSSSPS 828
QY 838 TTPNNMADPSSSTWSSSNEKPEITDNDVTAAPSLTPVPSAGOLRORSAPFPATATGSSPS 897
DB 829 TSNMNMADPSSSTWPTNTDQVSDNDVTAAPSLTPVPSAGOLRORSAPFPATATGSSPS 888
QY 898 PVLGGEKEVGLQAOALYPMRACKDHLNFKNSDVTTLVLEQDMMWFGVEYQGGKMPFYS 957
DB 889 PVLGGEKEVGLQAOALYPMRACKDHLNFKNSDVTTLVLEQDMMWFGVEYQGGKMPFYS 948
QY 958 YVKLISGPFRKSTSIDTPTSPASLKRVAAPAKPAIGEETIATYESSRQGLTLQ 1017
DB 949 YVKLISGPFRKSTSIDTPTSPASLKRVAAPAKPAIGEETIATYESSRQGLTLQ 1008
QY 1018 QGDVIVITKKGDMGTGVDKSGVPSNVRLKDSGSGTAKTSLGKKPEIAQVIA 1077
DB 1009 QGDVIVITKKGDMGTGVDKSGVPSNVRLKDSGSGTAKTSLGKKPEIAQVIA 1068
QY 1078 YVATGEQULTAPGQILIRKKNPGGMBGELQARKGQIGFPPANNYKLSPGTSKIT 1137
DB 1069 YVATGEQULTAPGQILIRKKNPGGMBGELQARKGQIGFPPANNYKLSPGTSKIT 1128
QY 1138 PTELPPTAVQPAVCOVIGWYDTAONDDELAFSGQIIVLNKEDDDMMKGEVSGVGF 1197
DB 1129 PTELPPTAVQPAVCOVIGWYDTAONDDELAFSGQIIVLNKEDDDMMKGEVSGVGF 1188
QY 1198 PSNYVLTMDPSQ 1213
DB 1189 PSNYVLTMDPSQ 1204

RESULT 5
ID ITN2_HUMAN STANDARD; PRT: 1696 AA.
AC Q9N2M3; Q95062; Q15812; Q9HAK4; Q9NKE6; Q9NYG0; Q9NZM2; Q9ULG4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Intersectin 2 (SH3 domain-containing protein.1B) (SH3P18) (SH3P18-like
DE WASP associated protein).
GN ITSN2 OR SH3D1B OR SWAP OR KIAA1256.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain;
RC TISSUE=Brain;
RX MEDLINE=200302722; PubMed=10922467;
RA Pucharcos C., Estivill X., de la Luna S.;
RT "Intersectin 2, a new multimodular protein involved in clathrin-
RT mediated endocytosis."
RL FEBS Lett. 478:43-51(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Seifert M., Engel M., Welter C.;
RT "Intersectin 2 (SH3D1B), human homolog of mouse Eps2 protein."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA TISSUE=Brain;
RC TISSUE=Brain;
RX MEDLINE=200309619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirokawa M., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:337-345(1999).
RN [4]
RP SEQUENCE OF 258-720 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 1261-1696

```

```

RP FROM N.A.
RC TISSUE=Embryo;
RA Iisgai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Makamata A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 821-1284 FROM N.A. (ISOFORM 1).
RA Wilson L.A., Fields D., Cruz L., Friesen J., Simionovitch K.A.;
RT "SH3P18-like wasp associated protein (SWAP): A multiple SH3 domain
RT containing protein that interacts with WASP."
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE OF 944-1191 FROM N.A. (ISOFORM 4).
RC TISSUE=Bone marrow;
RX MEDLINE=98294438; PubMed=9630982;
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
RT "Cloning of 139 and targets: systematic isolation of SH3 domain-
RT containing proteins."
RL Nat. Biotechnol. 14:741-744(1996).
RN [7]
RP ALTERNATIVE SPLICING.
RC TISSUE=Brain, and Fetal liver;
RX MEDLINE=21548828; PubMed=11690630;
RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;
RT "The human intersectin genes and their spliced variants are
RT differentially expressed."
RL Biochim. Biophys. Acta 1521:1-11(2001).
CC - FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC - SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
CC ITSN2 and Eps15, and different partners according to the step in
CC the endocytic process.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=ITSN2-L;
CC IsoId=Q9NZM3-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NZM3-2; Sequence=VSP_003893;
CC Name=3; Synonyms=ITSN2-S1;
CC IsoId=Q9NZM3-3; Sequence=VSP_003893, VSP_003894;
CC Name=4; Synonyms=ITSN2-S2, SH3P18;
CC IsoId=Q9NZM3-4; Sequence=VSP_003895;
CC - TISSUE SPECIFICITY: Ubiquitous. Isoform 1 is primarily expressed
CC in adult heart and liver.
CC - MISCELLANEOUS: Overexpression results in the inhibition of the
CC transferrin uptake and the blockage of the clathrin-mediated
CC endocytosis.
CC - SIMILARITY: Contains 1 C2 domain.
CC - SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC - SIMILARITY: Contains 1 EF-hand calcium-binding domain.
CC - SIMILARITY: Contains 1 PH domain.
CC - SIMILARITY: Contains 5 SH3 domains.
CC - SIMILARITY: Contains 2 EH domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage for and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF182198; AAF59903.1; -
CC EMBL; AF182199; AAF59904.1; -
CC EMBL; AF248540; AAF63600.1; ALT_INIT.
CC EMBL; AB033082; BAA66570.1; ALT_INIT.

```

DR EMBL; AK021545; BAB13841.1; -
 DR EMBL; AK000302; BA91068.1; -
 DR EMBL; AF001630; AAD00899.1; -
 DR EMBL; U61167; AAC50593.1; -
 DR HSSP; P29355; 1SEM.
 DR Genew; HGNC:6184; ITSN2.
 DR MIM; 604464; -
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_homology.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; RhogEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00036; ehand; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; RhogEF; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; RhogEF; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2 DOMAIN 2; 1.
 DR PROSITE; PS50010; DH 2; 1.
 DR PROSITE; PS00018; EF-hand; 1.
 DR PROSITE; PS50031; EH; 2.
 DR PROSITE; PS50003; PH DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 DR EndoCyclo; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 KM
 FT DOMAIN 21 109 EH 1.
 FT CA BIND 66 78 EF-HAND (POTENTIAL).
 FT DOMAIN 243 332 EH 2.
 FT DOMAIN 360 755 COILED COIL (POTENTIAL).
 FT DOMAIN 756 817 SH3 1.
 FT DOMAIN 897 955 SH3 2.
 FT DOMAIN 980 1038 SH3 3.
 FT DOMAIN 1052 1116 SH3 4.
 FT DOMAIN 1126 1185 SH3 5.
 FT DOMAIN 1208 1394 DH.
 FT DOMAIN 1433 1543 PH.
 FT DOMAIN 1555 1651 C2 DOMAIN.
 FT VARSPLIC 621 647 Missing (in isoform 2).
 FT VARSPLIC 1234 1248 Missing (in isoform 3).
 FT VARSPLIC 1249 1696 Missing (in isoform 3).
 FT VARSPLIC 1192 1696 Missing (in isoform 4).
 FT CONFLICT 290 290 V->I (IN REF. 4).
 FT CONFLICT 678 678 R->G (IN REF. 4).
 FT CONFLICT 827 827 KAVSPK->FAAAT (IN REF. 5).
 FT CONFLICT 944 950 WEPKSYV->EFAAAT (IN REF. 6).
 FT CONFLICT 1278 1284 GEKMPVO->VDAANS (IN REF. 5).
 FT CONFLICT 1552 1552 K->Q (IN REF. 2 AND 3).
 SQ SEQUENCE 1696 AA; 193329 MW; 17B5C8629BECF9B CRC64;

Query Match 48.9%; Score 3068; DB 1; Length 1696;
 Best Local Similarity 51.0%; Pred. No. 3.8e-110;
 Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;

Qy 1 MAQPTTPEGGLDVAITVERBAKDDQFLSKIPAGITTDQARNFFGGLPQVLAQ 60
 Db 1 MAQPTTANNGGPNWVAITSEERTKDRQFDNLKPSGYITTDQARNFPLQSGLPAPVLA 60
 Qy 61 IWLADNMNDGRMDQVERFSIMKLIKLIKLGQYQLPSTLPVWKQDPV--ALSSAPAFGIG 118

Db 61 IWLADNMNDGRMDQVERFSIMKLIKLIKLGQYQLPSTLPVWKQDPV--ALSSAPAFGIG 118
 Qy 119 GIASMPPLTAVAPVPMG-----SIPVNGSPPLVSVPPAAVPLANGAPV 165
 Db 120 ---SMPNISIPQPLPAPAPITSLSSATSGTNLPPLMPPTLPVPSVTSISLP---NGTASL 173
 Qy 166 IQPLPAHAHAATPKSSSFS--RSGGSGQNTKLGQAQS--FDVASAP----- 211
 Db 174 IQPLP-IPYSSSTIPHSSSYSLMMGFG--GASIQAAQSILDGSSSSSTSLSSGNS 229
 Qy 212 ---AAEMAVPQSSRLKYROLFNSHDXMSGHLTGPOARTILMQSSLPOAGLASIMNLS 267
 Db 230 PKGTSEMAVQPTRLKYRQKFNLTDXMSGYISGFQARNALLQSNLSQQLATITMLAD 289
 Qy 268 IDQGKLTAEFFILAMHLDVANSGLPPLVPEYTPPSRRARRSSGSSVSSSDVQ 327
 Db 290 VDGSGQLKAEFFILAMHLDMAKAGPLPLTLPPELVPPSFR---GKQI---DSING 341
 Qy 328 RLPEPSEDEQPEKPLPTFEDKKRNFERSGVELKGRQALLDQGRKQERLAQLER 387
 Db 342 TLPSYQMG--EEFQKLPPTFEDKKRANYERKGMELKRRQALMEQGRKAEKAKQEK 400
 Qy 388 AEQERKERQEQAKQOLEKQLEKQRELERQREERKEIERBAKRELERQOLE 447
 Db 401 EEMERKQRELEQEWKQOLEKRELERQREERERDIERBAKQELERORLE 460
 Qy 448 WERNROBLNQRKQEGGVYLKARKITLFELEALNDKKHQLQEGLODIRCLATQRO 507
 Db 461 WERIRQBLNQRKQREOEIVRLNSKKNLHLEALNGHQQISGLQVRLKQYQKT 520
 Qy 508 EIESTNKSRELRITAEITHLQOQLOESQOMGLRLEPKQILSDLOKQV--ONSLSRDSILT 566
 Db 521 ELEVLQDCDEIWEIQLQOELQOELQYONKLIYLPQOLANERIKNNQFSNTPDSGSL 580
 Qy 567 LKRALKEKELARQOLREQLDEVERETRSKLOEIDVFNNQ----- 605
 Db 581 HKSLKEKELC-QRLKQDLALBKETASKSEMDSFNNQKCGMDDSVLQCLLSLSCL 639
 Qy 606 -----LKELEIHSKQOLQKSLKARLQKQEKRSLELEKQ---EDAQRQVER 655
 Db 640 NNLFLLEKREITNTYNTQALAEQLYKIKRDKLEIKERKLELMQKLEDEAKAKAQK 699
 Qy 656 DKQLEHVQEQEQRPRKPHBEDLTKREDSVRKKEABERAKPEMQDQKSLFHHQEPAK 715
 Db 700 ENLWKENLKRKEE-----EKQKLEQEKQEKQEBERKAEKQKQKDXD----- 743
 Qy 716 LATQAPWSTTEKGLTISAQSVK-----VYYRALYFPESRSHDEITTIQPGDIVNDES 770
 Db 744 -----TIKAEKKKRETSVLVNYRALVPFARNDHDEMSFNSGDIIOQDEK 788
 Qy 771 QTGEPRGLGELKCKTGMPFANTAEKIP--ENEVPTAKPVTDLTSPAPKLAIRETPAP 828
 Db 789 TVGPRGLYGSFGQNFQMFPCNYVEKMPSENE-----KAVSPKAL----- 830
 Qy 829 LPVTSSPSTTPNNMAQFSSSTWPSSEKPETDWDTMAQPSLTVSAGOLRORSAFTP 888
 Db 831 LPPTVLSATS-----TSSEPLSSNQPASVTDYQ--VSPSNLTVTSMQ--KKSAPTR 880
 Qy 889 ATATGSSPSPYLGGEKVEYGLQALYPMRAKDNHNLFNKSDVITYLEQDMMWFEVQ 948
 Db 881 TVSPG-SVSPRHGGQVVENLKQALCSWAKKONHNLNFSGHDITYLEQDENMWFEVH 939
 Qy 949 GQKGMFPKSYVKLISGVRKSTSIDTPTSPSLKVAS--PAKALPGEETIANYTY 1006
 Db 940 GGRGMFPKSYVKLIPGSEVK-----REEPALVAANVKKPSTAAYSVEETIAYPY 991
 Qy 1007 ESSQGLTQGGQVIVYTKKDGDMWTGTVGDSGVPSNYVRLKDSGSGTACKTSG 1066
 Db 992 SSVEPGDLTTEGSEILVTDQDGMWGTSGIDRSGITPSNVKPKDDSPSASKSGASN 1051
 Qy 1067 KKPEIAQVIAVYATGEOQLTLAPGQILIRKKNPQGMWEGEGLQARKKQIQIGFPANVY 1126

Db 1052 KPPEIAQVTSAYVAGSSEQLSLAPQGLILKNTSGMNGEIQARCKRKQGWEPASHV 1111
 QY 1127 KLLSPGTSKTPTELPKTAQVAVGVGYDYTAQNDLELPSKQIITVINKKEPDW 1186
 Db 1112 KLLGSSSEATPAFH-----VCCVIMYDPAANNEDELSKQILINVKKDPDW 1164
 QY 1187 KGEVSGQVGLFPSNVKLTITDMDPSQ 1213
 Db 1165 QGEINGVTGLFPSNVYKMTTSDPSQ 1191

RESULT 6
 ITN2 MOUSE STANDARD; PRT; 1658 AA.
 AC 0920R6; 0920R5;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Intersectin 2 (SH3 domain-containing protein 1B) (EH and SH3 domains
 protein 2) (EH domain and SH3 domain regulator of endocytosis 2).
 GN ITSN2 OR SH3D1B OR ESE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RX MEDLINE=99164083; PubMed=10064583;
 RA Sengar A.S., Wang W., Bishay J., Cohen S., Egan S.E.;
 RT "The EH and SH3 domain Ese proteins regulate endocytosis by linking to
 dynam and Eps15.";
 RT EMO J. 18:1159-1171(1999).
 CC -1- FUNCTION: Adapter protein that may provide indirect link between
 the endocytic membrane traffic and the actin assembly machinery.
 CC May regulate the formation of clathrin-coated vesicles.
 CC -1- SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
 ITSN2 and Eps15, and different partners according to the step in
 the endocytic process.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Ese2L, Long;
 CC IsoId=Q920R6-1; Sequence=Displayed;
 CC Name=2; Synonyms=Ese2, Short;
 CC IsoId=Q920R6-2; Sequence=VSP_003896, VSP_003897;
 CC -1- TISSUE SPECIFICITY: Widely expressed in adult tissues.
 CC -1- DEVELOPMENTAL STAGE: Widely distributed throughout the adult
 forebrain. Prominent expression was observed in the neocortex, the
 piriform cortex, the pyramidal cell layers of hippocampus, the
 dentate gyrus, in several nuclei of the thalamus and hypothalamus
 and in the amygdala.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
 CC -1- SIMILARITY: Contains 1 EF-hand calcium-binding domain.
 CC -1- SIMILARITY: Contains 1 PH domain.
 CC -1- SIMILARITY: Contains 5 SH3 domains.
 CC -1- SIMILARITY: Contains 2 EH domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF132479; AAD19747.1; -;
 DR EMBL; AF132480; AAD19748.1; -;
 DR HSSP; P29355; 1SEM.
 DR MGI; MGI:138049; Sh3d1b.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_homology.

DR InterPro; IPR000108; Neu_cyt_fact_2.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR000219; Rhogef.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00168; C2; 1.
 DR Pfam; PF00036; eefhand; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00621; Rhogef; 1.
 DR Pfam; PF00018; SH3; 5.
 DR PRINTS; PR00459; P67PHOX.
 DR ProDom; PD000066; SH3; 5.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00325; Rhogef; 1.
 DR SMART; SM00326; SH3; 5.
 DR PROSITE; PS50004; C2 DOMAIN_2; 1.
 DR PROSITE; PS50010; DH_2; 1.
 DR PROSITE; PS00018; EF_HAND; 1.
 DR PROSITE; PS50031; EH_2.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50002; SH3; 5.
 KW Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
 KW Alternative splicing.
 FT DOMAIN 21 109
 FT CA BIND 66 78
 FT DOMAIN 244 333
 FT DOMAIN 364 716
 FT DOMAIN 717 778
 FT DOMAIN 851 909
 FT DOMAIN 941 999
 FT DOMAIN 1013 1077
 FT DOMAIN 1087 1146
 FT DOMAIN 1169 1356
 FT DOMAIN 1395 1505
 FT DOMAIN 1517 1613
 FT VARSPLIC 1187 1197
 FT FTID=VSP_003896.
 FT VARSPLIC 1198 1658
 FT CONFLICT 451 452
 FT CONFLICT 1161 1161
 FT SEQUENCE 1658 AA; 188776 MW; 7050EFC5F7983A5 CRC64;
 SQ
 Query Match 48.6%; Score 3049.5; DB 1; Length 1658;
 Best Local Similarity 52.4%; Pred. No. 1.9e-109;
 Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
 QY 1 MAQFTPPGSLDVAITVEERAKHDQFLSKPIAGFITGDQANFFPQGLPQVLAQ 60
 Db 1 MAQFTPANQGPNNWAIITSEERTGDKQFDNFKSGGYITGDQARTFLQGLPAPVLA 60
 QY 61 IWAADNNNDNRMOVERSIMKLIKELQGYVQVPTLPPMKQPV--AASSAPAFGIG 118
 Db 61 IWAASLDKDKMDQDEFSIMAKLIKELQGYVQVPTLPPMKQPV--AASSAPAFGIG 119
 QY 119 GIASNP-----PLTAVAPV--PMG-----SIPVGSPLVSSVPPAAVPLANGAP 164
 Db 120 ---SNPNSTHQPLPVPVAPVATPLSSATSGTSIPPLMAPLVBSVTSILP---NGTAS 173
 QY 165 VIQPLPAFPAHPAATWPKSSFS--RSGGSQLNTKLAQAS--PDVASAP----- 211
 Db 174 LIQPL-SIPYSSSTLPHASSYSILMMGGRG---GASIQAQSLIDLGSSSSSTSTASLGN 229
 QY 212 -----AAFWAPQSSRLKYROLFNSHDTWSGHITGPARTIMOSSIPQAOIASIMVLS 266
 Db 230 SPKTGTSMAVAPQPSRLKTRQKFNSLDKMGVYISGFAVALAQSNLSQTQLATWTLA 289
 QY 267 DIDDGKLTAAEFILAMLIDVAMSGQPLPVLPPEYVPPSFRRRSSGSGVSISSSSVD 326
 Db 290 DIDGGQAKAEFILAMLTMAKAGQPLPLTLPPELVPPSFR-----GKQV---DSVN 341

```

327 ORLEBESSBDEQPEKLPYTFEDKKRNERGSELEKRRQALLLEQQRKEERLQLE 386
342 GTLSPSYQTQ-EEBPOKCLPVTFEDKRYANERGMWLEKRYOVMLEQQRKEERKQKE 400
387 RAEQERKERQEOBQAKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 446
401 KEHEERKQRELOEQBQKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQ 460
447 EWEENRQELINQNRKEQGVTVLKARKKTELEFELALNDKKGHLEGLQDIDRCLATQ 506
461 EWERLRQELLSQKREQEDIVRLSSRKKSJLLELVANGKHQISGRLOVQIRKQYQK 520
507 QEISTNSRRLRLAETHLOOQLOESQOMLGRLIPKQILSDQKQVQNSLHRDLSLT 566
521 TELEVLDDQCDLETHIKOLOELKRYQNKLIYLPKQILNRIKNNQNTSDPSGL 580
567 LKRALAEKELARQQLREQLDEVERETRSKLOEIDVFNQLEKELREHKSQOLQOKRSLEA 626
581 LHKSSSEKELCQRLKEQDLALEKKTAKLSGEMDSFNNQLEKELRESYVTOQLALEQLHI 640
627 ARLKQKEQKSLF-LEKQK--EDAQRRVQERDKQMLEHVQOEQPRPRKPHEDRLKR 682
641 KRDKLKEIERKRLQIQKKLEDEAKRQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 694
663 EDSVRKKEAERAKEMQDKQSLFHPHQEPAKLATQAPMSTTEKGLTISAQESVKYV 742
695 EKSQDKTEER--KAERKQSE-----TAAH-----LVN 721
743 YRALYPFSSRSHDEITIQGDIVWVDESQGEPCWGLGELKQKGMFPYAEKIPENEV 802
722 YRALYPFARHNDHESFSSGDIQVDEKTEVEPGWLYSPQCKPFQPCVNYEKLSE- 780
803 PTPAKPVTDLTAPAKLALRETPAPLPVTSESEPTTNNMADFSTPSSSNEKPEFDN 862
781 -----KALSPKXALLPPTVSLSTATSTS-SQPPASVTDYHNV--SFSNLTVNT-- 824
863 WDTMAAQSLTVPSAGQURQSPAFPATATSSPSPVQGGKVGGLQAOALPYPARAKD 922
825 --TW-----QOKSAFRTVSPG-SVSPHGGQGVANLQALCSWTAKKE 867
923 NNLNFKSDVITVLEEQDMMWFGVEYQKGMFPKPSYKLSG-PYRKS--TSIDTPTES 979
868 NNLNFKSDVITVLEEQDMMWFGVEYHGRGMPKPSYKLIPENEQDRBPALVAAVTKK 927
980 PASLRVASPAKPAIP-GEFTIAMYTESSEQGLTFOQGDVIVYTKKQDGMWGTGVD 1038
928 PTS--TAYPVSTAYPVGEDYIALYSYSVPEPGDLTFEGEELIVTQKQDGMWGTGIGE 984
1039 KSGVPSNRYVRKQSESGTAGKTSQKGPRIAGVIVASVATGEOLTLAPGQILIRK 1098
985 RTGIPSPNRYVRKQDENFGNASKSASNNKGPRIAGVIVASVATGEOLTLAPGQILIRK 1044
1099 KNPQGMWGELOARQKQKQIGWFPANVYVLLSPGTSKITPTTELPTAVQAPVQVIGMYD 1158
1045 KNTSGMGELOARQKQKQKGMFPASHVYVLLGSPSERMPT-----FHANCOYIANWD 1097
1159 YTAQNDDELAFSKGQIINVLANKEDDPMWKGVEVSGOGLFPSPNRYVYKLTDMDSQ 1213
1098 YWANNEDDELAFSKQILINVMKDDPMWQGETNGTGLTFPSNRYVIMVTTSDPSQ 1152

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC NCBI_TaxID=9606;
CX [1]
RX SEQUENCE FROM N.A.
RP TISSUE=Melanoma;
RC MEDLINE=94239734; PubMed=8183552;
RA Wong M.T., Kraus M.H., Carlomagno F., Zelano A., Druck T.,
RA Croce C.M., Huebner K., di Fiore P.P.,
RT "The human eps15 gene, encoding a tyrosine kinase substrate, is
RT conserved in evolution and maps to 1p31-p32."
RL Oncogene 9:1591-1597(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94181254; PubMed=8134107;
RA Bernard O.A., Mauchaufe M., Mecucci C., van den Berghe H.,
RA Berger R.;
RT "A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not
RT related to AF-4, AF-9 nor ENL."
RL Oncogene 9:1039-1045(1994).
RN [3]
RP INTERACTION WITH CRK.
RX PubMed=8662907;
RA Matsuda M., Ota S., Tanimura R., Nakamura H., Matuoka K., Takenawa T.,
RA Nagashima K., Kurata T.;
RT "Interaction between the amino-terminal SH3 domain of CRK and its
RT natural target proteins."
RL J. Biol. Chem. 271:14468-14472(1996).
RN [4]
RP STRUCTURE BY NMR OF 217-311.
RX MEDLINE=20222955; PubMed=10757979;
RA Emmon J.L., de Beer T., Overduin M.;
RT "Solution structure of Eps15's third EH domain reveals coincident
RT Phe-17p and Asn-Pro-Phe binding sites."
RL Biochemistry 39:4309-4319(2000).
RN [5]
RP STRUCTURE BY NMR OF 121-215.
RX MEDLINE=98387926; PubMed=9721102;
RA de Beer T., Carter R.E., Lobel-Rice K.E., Sorokin A., Overduin M.;
RT "Structure and Asn-Pro-Phe binding pocket of the Eps15 homology
RT domain."
RL Science 281:1357-1360(1998).
CC -1- FUNCTION: Involved in cell growth regulation. May be involved in
CC the regulation of mitogenic signals and control of cell
CC proliferation. Involved in the internalization of ligand-inducible
CC receptors of the receptor tyrosine kinase (RTK) type. In
CC particular EGFR (By similarity).
CC -1- SUBUNIT: Interacts with Stc1n 2. Interacts with CRK via its SH3-
CC binding sites.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Recruited to the plasma
CC membrane upon EGFR activation and localized to coated pits (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed.
CC -1- PTM: Phosphorylation on Tyr-849 is involved in the internalization
CC of EGFR. Not required for membrane translocation after EGF
CC treatment or for targeting to coated pits, but essential for a
CC subsequent step in EGFR endocytosis (By similarity).
CC -1- DISEASE: Involved in a t(1;11)(p32;q23) chromosomal translocation
CC in acute leukemias causing fusion to the c-ABL gene (MLV or HRX)
CC gene product which contains DNA-binding motifs resulting in a
CC rogue activator protein.
CC -1- SIMILARITY: Contains 3 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 2 ubiquitin-interacting motif (UIM) repeat.
CC -1- DATABASE: NAME=Atlas Gene. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.intoblogen.fr/services/chromosome/Genes/AF1pD11.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
DR EMBL; U07707; AAA52101.1; -.
DR EMBL; Z29064; CAA82305.1; -.
DR PIR; S43074; S43074.
DR PDB; 1C07; 19-JUL-00.
DR PDB; 1EH2; 22-JUL-99.
DR PDB; 1F9H; 01-NOV-00.
DR PDB; 1F9I; 01-NOV-00.
DR Gene; HGNC:3419; EP815.
DR MIM; 600051; -.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0007173; P:EGF receptor signaling pathway; TAS.
DR GO; GO:0007048; P:oncogenesis; TAS.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00261; EPS15_homology.
DR InterPro; IPR003903; UIM.
DR Pfam; PF00036; efhand; 4.
DR SMART; SM00054; EFh; 3.
DR SMART; SM00027; EH; 3.
DR SMART; SM00726; UIM; 3.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS00031; EH; 3.
DR PROSITE; PS00330; UIM; 2.
DR SH3-binding; Phosphorylation; Calcium-binding; Repeat;
KW Chromosomal translocation; Proto-oncogene; 3D-structure.
FT DOMAIN 15 104 EH 1.
FT CA_BIND 128 216 EH 2.
FT CA_BIND 173 184 EF_HAND 1 (POTENTIAL).
FT CA_BIND 224 314 EH 3.
FT DOMAIN 236 247 EF_HAND 2 (POTENTIAL).
FT DOMAIN 599 827 EF_HAND 2 (POTENTIAL).
FT REPEAT 623 625 1.
FT REPEAT 629 631 2.
FT REPEAT 634 636 3.
FT REPEAT 640 642 4.
FT REPEAT 645 647 5.
FT REPEAT 651 653 6.
FT REPEAT 664 666 7.
FT REPEAT 672 674 8.
FT REPEAT 692 694 9.
FT REPEAT 709 711 10.
FT REPEAT 737 739 11.
FT REPEAT 798 800 12.
FT REPEAT 804 806 13.
FT REPEAT 825 827 14.
FT DOMAIN 851 870 UIM 1.
FT DOMAIN 877 896 UIM 2.
FT DOMAIN 768 850 PRO-RICH.
FT SITE 774 774 SH3-BINDING.
FT MOD_RES 849 849 PHOSPHORYLATION (BY EGFR) (BY
FT CONFLICT 822 822 SIMILARITY).
FT SEQUENCE 896 AA; 38673 MW; A1B9FB04A07FABEB CRC64;
Query Match 7.8%; Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 4,7e-12;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;
QY 15 MAITVERAKHDQOFLSLKPIAGFITGDQARNFFQSGLPQVLAQIMALADWNNDRMD 74
DB 122 MAVKPEADKAYDIFSLSPVNGFLSGDKYKPYVLANSKLPVLDLGRVWELSDIDHGMID 181
QY 75 QVEFSIAMKLIKLKIGYOLPSTLPPVMQOPVAISSAPAFGIGIASMPPLTAAPVPM 134
DB 182 RDEFAVAMFLVYCALR-----KEPVPV 203
QY 135 GSIPVVGMSPLVSSVPAPAVPPLANGAPPVIOPLPAFAHPATWPKSSFSKSGGSQL 194
DB 204 -----SLPPALVPP-----SKR 215
QY 195 NTQLQKAGSFDVASAPPAEAWAVQSSRLKYROLFNSHDKTMSGHLTGPOARTILMOSSL 254

```

```

DB 216 KT-----WVSPAKAKYDIEIFLKTDXMDMGFSGJEVREIFLKTGL 257
QY 255 POAQLASIMNLSIDIDODGKLTAEFLIAMLHLDVAM-SQOPLPVLPPEYIPPSFRVNS 313
DB 258 PSTLAAHWSLCDTDKDCGLSKDQFALAFHLISOKLIKIDIPPHVLTPEMIPPS---DFA 314
QY 314 GSGMSVYSSSVYDQRLPEEPPSSDEQPEPKLPVPEDEKKRNFERSGVLEKRRQALLE 373
DB 315 SLQKNTIGSSPV-----ADPSAIK-----ELDTLNEIYD 344
QY 374 QQRKEQERLAOLERAEOEKREKEROEAKROLELEKOLEKOREROREEERKEIERR 433
DB 345 LQREK-----NNVQDLKEK-----DTIKQRTSVQ---DLQDEVQHE 380
QY 434 EAAKRELEROROLEMERNRROSELNDORNKQEGESTVVLKARKRTLEBELANDKQKQLG 493
DB 381 NTNLOKLOAKOQ-----QVEELIDE-----LDEQKQLEB 410
QY 494 KLQDIRCRATROEIESTNKSRELRIAEITHLQOOLQESQOMAGRIPEKQILSQOLQ 553
DB 411 QLEKVAKKCAEHAQLISSLK-----AELTQESQISTYEBELAK-----ARELSR 456
QY 554 VOQNSLHRDSLTLTKRALBAKELARQOLREOLDEVERETRS-----KLQEIDVFNNOUK 607
DB 457 LQQETAE-----LESVSVSKAQLEPLQOQHQDSQQELISSMQMKLMEKDLLENHNSQLN 510
QY 608 ELREIHS-----KQOLQKORSLEAPARKQEKQERSLEL--- 641
DB 511 WCSSPHSILVNGATDYCSLSTSSSEPNANLNEHVEGCSNLESEPIHOESPARSSPELLPSG 570
QY 642 EKQKEDAGORVERDQOMLEHVQOEQRPRPKRPHEDRLKREDSVAKKAEBR----- 644
DB 571 VTDENEVTTAVTEKQCELDN-----NRSKEDPPNNVSSSLTGPAVDNLDPFGS 622
QY 695 -----AKPEMOKOSRLFFHPOEPAK-----LATQAPWSTTEKGLTISAOESV 738
DB 623 DPFVSGDPFKDPPFGKIDPFGDPPFGSDPPFASDCFRSTDPFAISSDTPFSAANNSI 682
QY 739 KVVYTRALYPFESRSHDEITIQGDIWMVDESQT-----GEPMGLG----- 780
DB 683 TSV-----ETLKINDPFAFGVTVAAASDATDPFASVFQNESGCGFADFSTLSKYV 734
QY 781 -----ELKGTGWFPAN-----YAEKIPENEVTPAKPVTDLTSAPAKLARETPAPLP 830
DB 735 NEDPFESATSSSVSNVITKNVFETSVKSEDEPPALP-----PRIGTPTRCPPLP 785
QY 831 -----VTSSEP-----STTPNNMADFSTWPSSENEKPEETDNWPT 865
DB 786 PKKRSINKLDSDFPRLANDPQFPFGNDSPKREKDPKCDPFTSATTTTNKENDPSPNAN 845
QY 866 WAAQPS 871
DB 846 FSAVPS 851
RESULT 8
EPI5_MOUSE
ID EPI5_MOUSE STANDARD; PRT; 897 AA.
AC P42567;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Epidermal growth factor receptor substrate 15 (Protein Eps15) (AF-1p
DE protein).
GN EPS15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=fibroblast;

```

RX MEDLINE=93361014; PubMed=7689153;
 RA Paoletti F., Minichiello L., Matochkova B., Wong M.T., di Fiore P.P.;
 RT "Eps15, a novel tyrosine kinase substrate, exhibits transforming
 RT activity.";
 RL Mol. Cell. Biol. 13:5814-5828(1993).
 RN (2)
 RP PHOSPHORYLATION OF TYR-850, AND MUTAGENESIS OF TYR-850.
 RX MEDLINE=20411232; PubMed=10953014;
 RA Confalonieri S., Salcini A.E., Puri C., Tacchetti C., Di Fiore P.P.;
 RT "Tyrosine phosphorylation of Eps15 is required for ligand-regulated,
 RT but not constitutive, endocytosis.";
 RL J. Cell Biol. 150:905-912(2000).
 RN [3]
 RP STRUCTURE BY NMR OF 7-105.
 RX MEDLINE=99400466; PubMed=10471276;
 RA Whitehead B., Tessari M., Carotenuto A.,
 RA van Bergen en Henegouwen P.M., Vulliamis G.W.;
 RT "The Eps15 domain of Eps15 is structurally classified as a member of
 RT the SH3 domain class of EF-hand-containing proteins.";
 RL Biochemistry 38:11271-11277(1999).
 CC -1- FUNCTION: Involved in cell growth regulation. May be involved in
 CC the regulation of mitogenic signals and control of cell
 CC proliferation. Involved in the internalization of ligand-inducible
 CC receptors of the receptor tyrosine kinase (RTK) type, in
 CC particular EGFR.
 CC -1- SUBUNIT: Interacts with Crk via its SH3-binding sites (by
 CC similarity). Interacts with Srcin 2.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Recruited to the plasma
 CC membrane upon EGFR activation and localized to coated pits.
 CC -1- PTM: Phosphorylation on Tyr-850 is involved in the internalization
 CC of EGFR. Not required for membrane translocation after EGF
 CC treatment or for targeting to coated pits, but essential for a
 CC subsequent step in EGFR endocytosis.
 CC -1- SIMILARITY: Contains 3 EF domains.
 CC -1- SIMILARITY: Contains 2 ubiquitin-interacting motif (UIM) repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; L21768; AAA02912.1; -
 DR PIR; A54696; A54696.
 DR PDB; 1KYF; 12-JUN-02.
 DR PDB; 1KYU; 12-JUN-02.
 DR PDB; 1QUT; 11-SEP-02.
 DR MCD; MGI:104583; Eps15.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; Eps15_homology.
 DR InterPro; IPR003903; UIM.
 DR Pfam; PF00036; ehand; 4.
 DR SMART; SM00054; Efh; 4.
 DR SMART; SM00027; EH; 3.
 DR SMART; SM00726; UIM; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS50031; EH; 3.
 DR PROSITE; PS50330; UIM; 2.
 KW Phosphorylation; Calcium-binding; Repeat; 3D-structure.
 FT DOMAIN 15 104 EH 1.
 FT 128 216 EH 2.
 FT CA BIND 173 184 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 224 314 EH 3.
 FT CA BIND 236 247 EF-HAND 2 (POTENTIAL).
 FT DOMAIN 599 829 15 x 3 AA REPEATS OF D-P-F.
 FT REPEAT 599 601 1.
 FT REPEAT 623 625 2.
 FT REPEAT 629 631 3.
 FT REPEAT 634 636 4.
 FT REPEAT 640 642 5.

FT REPEAT 645 647 6.
 FT REPEAT 651 653 7.
 FT REPEAT 665 667 8.
 FT REPEAT 673 675 9.
 FT REPEAT 693 695 10.
 FT REPEAT 711 713 11.
 FT REPEAT 806 808 12.
 FT REPEAT 827 829 13.
 FT DOMAIN 852 871 UIM 1.
 FT DOMAIN 878 897 UIM 2.
 FT MOD_RSS 850 850 PHOSPHORYLATION (BY EGFR).
 FT MUTAGEN 850 850 Y->F: INEFFICIENT EGFR INTERNALIZATION.
 SQ SEQUENCE 897 AA; 98470 MW; 08A0CD0423F873C2 CRC64;
 Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.9e-12;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;
 15 WAITERAKHDDQFLKPIAGFITDQANFFQSGLPQPVLAQIWMALADNNDGRMD 74
 122 MAVSEDKAKYDAIFDLSIPVDGSLSDKYPVLNKLPIEILGRWELSDIDHDKLD 181
 75 QVEFSIAMKLIKLTQGYLPSTLPVWKQOPVAISSAPAFGIGIASMPPLVAVPYM 134
 182 RDEFVAMFLVYCALR-----KEPVP 203
 135 GSITVGMSPPLVSVPPAAPPVPLANGAPVIOPLPAFPAATWPKSSFSRSGPSQL 194
 204 -----SLPPALVP-----SKR 215
 195 NTKLQKASFPVAPASPAEAEMAVPQSSRLKYROLFNHSDKMTSGHTGPQARTILMOSL 254
 216 KT-----WVSPAKAKYDEFLTKTDMDGVYGLVRETFRTGL 257
 255 PQAGIASIMWISDIDODGKLTAEFFILAMLIDVAM-SCGPLPVLPEYIPSPFRVRS 313
 258 PSALLAHWSLCDTKGCKSLKSDQFALAFILINQKLIKIDIPSHLPMPSPS---DRS 314
 314 GSGSVSVSSSYDQRLPEEBSSEDEQPEKLPTEPEDKKRENERBSVLEKROALLE 373
 315 SLQKNITGSSPV-----ADFAIK-----ELDTLNNEIVD 344
 374 QORKEORLAQLAEORERKERERQOEAKKQLELEKQRELEROREERKEIER 433
 345 LQREK-----NNVEDLKE-----DYKQKRSVQ---DLQDEVORE 380
 434 EAARLEERQROLEWERNRRQELINQNKQEGTVLAKRKTLFELEALNDKKHQLQEG 493
 381 SINQKLOAQKQ-----QVQELLGE-----LDEQKQALEE 410
 494 KLQDIRCLATQROEISTNKSRELRLAETHTLQOQLOESQMLGRLLPEKQISDQLO 553
 411 QLOEVRKKCAEALISLK-----AETISQSSQISYEELTK-----AREELSR 456
 554 VQNSLHRDLSLTTRALAEKELARQOLREQLDEVERE-----TRSKLOEIDVFNNOKE 608
 457 LQQTAAQ-----LEESVESKQALPELQQLQSQOQISSMQRLMKQLJETNNQ--- 507
 609 LREIHSKQLOLQKQ-----RSLAARLQKQEKERSLELEKQEDAQRRVQ 653
 508 -----SNWSSSPQVLVNGATDYCSLSTSSSETANFNHAGQNNLSEPHQSSVRS 562
 654 ER-----DKQMLEHYQOEQPPRPKPHEDRLKRDYSVKKA-----ERAPENQDK 702
 563 PEIAPSDVTSEAVTVAQNEKVTPL--FDDDKRSKEEDPFNVSSSLTDAVADTNLIDF 620
 703 QSRLE-----HPHOEPAKLTAQAPWSTTEGPTLISAOESVKKVYV 743
 621 QSDPFVSGDPKDPFGKIDPFGDPPKGSPPFASDCEFKQISTDPFTSTSDPFSASN 680
 744 RALYPFESRSHDEITIGPDIWVDESQTEBP--GWLQGLKIGKTYGFPANVAKIDENE 801
 681 SSNTSVETWKNDPFAQGGTVVAASADSATDPFASVFGNEBFG--DGFADFTLSKVNED 739

QY 802 VETP-----AKPVTDLTSA-----PA--PKLALRENPAPLPTTSSEPTTPNN 842
 DB 740 ANNPRTSSSTSSVTAKPKMLBETASSEVPPALPKVGTTPRCPDP-----PGKRPIN 794
 QY 843 WDFSSST-----WSSSNKEPRTDNDMTMAAQPSLITVPSAGQLRQSAFTPATATGSS 895
 DB 795 KLDSSDPLKLNDFQFPQPNDSFKKXDPMMFC-----DPFTSSTTTNKE 838
 QY 896 PSP 898
 DB 839 ADP 841

RESULT 9
 PANI_YEAST STANDARD; PRT; 1480 AA.
 AC P32521;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE PANI protein.
 GN PANI OR MIP3 OR MDP3 OR YIR006C OR YIB6C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 320-344; 352-375 AND 899-906.
 RX MEDLINE=92405166; PubMed=1339314;
 RA Sachs A.B., Dearsdorff J.A.;
 RT "Translation initiation requires the PAB-dependent poly(A)
 RL Cell 70:961-973 (1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Timmes J., Teodora C., Valencia A., Sengen C., Wiemann S.,
 Schwager C., Zimmermann J., Sander C., Ansoyge W.;
 RT "Nucleotide sequence and analysis of the centromeric region of yeast
 RL Chromosome IX.";
 RN Yeast 11:61-76(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX PubMed=9169870;
 RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 RA Harris D.E., Hornsby T., Hunt S., Jagsels K., Jones M., Lye G.,
 RA Moulis S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrall B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
 RL Nature 387:84-87(1997).
 CC -!- FUNCTION: NOT KNOWN.
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.
 CC -!- SIMILARITY: Contains 2 EH domains.
 CC -!- CAUTION: WAS ORIGINALLY THOUGHT TO BE A SUBUNIT OF PAB-DEPENDENT
 CC POLY(A)-SPECIFIC RIBONUCLEASE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z38062; CAAB6208.1; -
 DR EMBL; X79743; CAAB8097.1; -
 DR EMBL; M90688; AAA34841.1; -
 DR PIR; S48440; S48440.
 DR SGD; S0001445; PANI.

DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomycetes); IDA.
 DR GO; GO:0005886; C:plasma membrane; IDA.
 DR GO; GO:0008093; F:cytoskeletal adaptor activity; IPI.
 DR GO; GO:0007120; P:axial budding; IMP.
 DR GO; GO:0016288; P:cytokinesis; IMP.
 DR GO; GO:0006897; P:endocytosis; IMP.
 DR GO; GO:0007121; P:polar budding; IMP.
 DR InterPro: IPR000261; EPR15_homology.
 DR Pfam: PF00036; ehband; 3.
 DR SMART; SM00054; EFh; 3.
 DR SMART; SM00027; EH; 2.
 DR PROSITE; PSS0031; EH; 2.
 KW Repeat.
 FT DOMAIN 142 575 15 X 12 AA APPROXIMATE REPEATS.
 FT REPEAT 142 153 1-1.
 FT REPEAT 164 175 1-2.
 FT REPEAT 188 199 1-3.
 FT REPEAT 215 226 1-4.
 FT REPEAT 235 246 1-5.
 FT DOMAIN 270 359 EH 1.
 FT REPEAT 352 403 1-6.
 FT REPEAT 409 420 1-7.
 FT REPEAT 422 433 1-8.
 FT REPEAT 446 457 1-9.
 FT REPEAT 467 478 1-10.
 FT REPEAT 488 509 1-11.
 FT REPEAT 507 518 1-12.
 FT REPEAT 538 549 1-13.
 FT REPEAT 545 556 1-14.
 FT REPEAT 564 575 1-15.
 FT DOMAIN 328 680 2 X 23 AA APPROXIMATE REPEATS.
 FT REPEAT 328 350 2-1.
 FT DOMAIN 600 689 EH 2.
 FT REPEAT 600 689 2-2.
 FT DOMAIN 1084 1125 7 X 6 AA TANDEM REPEATS.
 FT REPEAT 1084 1089 3-1.
 FT REPEAT 1090 1095 3-2.
 FT REPEAT 1096 1101 3-3.
 FT REPEAT 1102 1107 3-4.
 FT REPEAT 1110 1113 3-5.
 FT REPEAT 1114 1119 3-6.
 FT REPEAT 1120 1125 3-7.
 FT DOMAIN 1315 1377 8 X 6 AA APPROXIMATE REPEATS.
 FT REPEAT 1315 1320 4-1.
 FT REPEAT 1321 1326 4-2.
 FT REPEAT 1327 1332 4-3.
 FT REPEAT 1340 1345 4-4.
 FT REPEAT 1345 1350 4-5.
 FT REPEAT 1355 1360 4-6.
 FT REPEAT 1361 1366 4-7.
 FT REPEAT 1372 1377 4-8.
 FT DOMAIN 13 22 POLY-GLN.
 FT DOMAIN 29 34 POLY-GLN.
 FT DOMAIN 98 106 POLY-GLN.
 FT DOMAIN 1400 1406 POLY-PRO.
 FT DOMAIN 1452 1455 POLY-GLU.
 FT DOMAIN 1474 1480 POLY-PRO.
 FT CONFLICT 265 235
 FT CONFLICT 266 273
 FT CONFLICT 474 487
 FT CONFLICT 653 657
 FT CONFLICT 1291 1291
 FT CONFLICT 1396 1480

SO SEQUENCE 1480 AA; 160267 MW; F3518455FF59553 CRC64;
 Query Match 6.5%; Score 410.5; DB 1; Length 1480;
 Best Local Similarity 20.5%; Pred. No. 6.1e-09;
 Matches 261; Conservative 174; Mismatches 444; Indels 395; Gaps 47;


```

OY 17 ITVERAKHQDOQLSKPI-AGFITGQDARNFPPFSGSLPOPVLAIQIYALADMNNDGRDQ 75
Db 266 ITAODQAKFETLFSIVTNGSNTVSGANCKILMRSGSLPQSOLARITWLCITSAAGELLF 325
OY 76 VEFISAMKLIKLIKQYQPLSTLPPVMKQ----- 104
Db 326 PEFALAMHLINDVLOQDITPYELDSKTKNEVSSFIDAINLSIANQDSSANDAPKTPDEF 385
OY 105 -----QPVASAPAFGI-----GGTASM----- 123
Db 386 ITAGVONLQOPQPTGYMPTSGIFLQSQITGGVASALNPOSTGFMATPTENMSNTGTP 445
OY 124 ---PPLTAVAVPM-----GSIP-VYGMSPPLVSSVPPAAV-----PPLAN 160
Db 446 GLNPQITGAPASMPNITGNALQPTTGMMPQTTGMMPQTTGMMPQTSFGVNLGFLTG 505
OY 161 GAPVITQP-----LPFAHAPATWPKSSFSRSGPSQL----- 194
Db 506 GA---LQSQYTGYSVMPQSGPASM---PNLSFNOQLQSQITGLQPTGFLPPSNFS 560
OY 195 -----NTKQKQASFDVASAPPAEAMVPPSSRLKTRQLNSHDKTMSGLT 241
Db 561 ATMPLTACKTGFNGNEIYTKSNFNNLIDNSSQDKISTEEKSLFYKIFETDTQNGKLLD 620
OY 242 GPOARTILMOSSLPOAQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGPLPPVLP 301
Db 621 SPTAVEIRKSGILRADLEQIMNLCIDINNTQNLNQBALMHLVYGLTKNGKPIPNVLP 680
OY 302 EYIP-----DSF-----RVRSG- 314
Db 681 SLIPSTKLNLNKLKQLTPTTTEKESFGKIDALSYKNDDVLPVYRRRKVYSAKN 740
OY 315 ---SGMSIVSSSVUDQRLPEBSSSDEQPEKKLPVTEDEK---KRNFERGSVELEKR 367
Db 741 EEOSSFSPPSAKSVNH-----SSSTLTQTDISVKTVEKTKAKVYAGFR---EIMLK 791
OY 368 ROALLEQORKEQER-----LAQLEAEQERKEERQOEAKROLE 407
Db 792 NIASLENIKINISNENCYDSSIPSDLTSRPAITAKPMLFNEISTIDNETAKKIDLY 851
OY 408 LEKQLEKQERLEERERREKREIEREAKRELEROROLEWERNRQELNQRKEQECT 467
Db 852 RKK---NPSIISGSPNGEITENDRKAKSRALLR-----ARMALTGKSTESBDS 899
OY 468 VILKARKRTILEFE-LEALNDKQHOLEGLQDRCRLATVOROEIESTNSRELRITLHL 526
Db 900 LSMEDEQOSAEIKRIQOENGNQOI---IKDIRSSISDISASLAKSTMTGSMN---ISN 951
OY 527 QOQLOESQOMLGRLLPEKQILSDQKQVQONSIAHDS----- 563
Db 952 -QEFERMEFGIGLEQVAREFL-DLJKSNSKSVTSSFPVBSSTTPVDDSSSPSYGOF 1009
OY 564 LITLRALEKELARQOLREQLDEVERETRSKLQSIDVFN-NQKELREIHSKQOL--- 618
Db 1010 KTAERAAVYLKEQAKKMEKELAKFDKRNRYTQSSRSISSNSREQOQLAGSSNLVBP 1069
OY 619 -----QKRSLEAALKQEQERKSLLEKQK----- 646
Db 1070 RATPPEQEKYEVAPQTPVOSTQVQPTQVQPTQVQPTQVQPTQVQPTQVQVNY 1129
OY 647 DAQRVQERDKQWLHVOOEQPRPKPHEEDRLR-----BDSVRKKEAE----- 693
Db 1130 NAKQSDDEDEDEDEKRLQEBELKRLKLLKKADKEKGLAALRQIIDAQNESBEBETNGKD 1189
OY 694 -----RAKP-----EMODKQSLFPHQEPA---KLATQAPWSTTEKGPLTISA 734
Db 1190 NFGGVHNVQAPAVAPSAAFQNSNTAPRSVAATVPAAGKSTGLPSTTMGHNVYFQDA 1249
OY 735 QESVAVYVYRALYPRESSHD-EITIQEDIVMVBESQTEGEGWIGELKGTGTFPANY 793
Db 1250 SAS-----STSTPDARAEMQRIQR-----LDEDE---DDGWSDEDESNNR---VAV 1292
OY 794 AEKIPENEVPTP---AKPVDLTSAAPAKLALRETPAP---LPVTSSEPTTPNNWADFS 847

```

```

Db 1293 DNKYEAKIGHPDRAPV---TAAPLPSV---TVPVAVPVPQANTSNKSSPIPIA 1345
OY 848 STWSSSNE-----KEPTDMWDTPAOPSLTVSAGLRQRS-AFTPATATGSSPSVLG 901
Db 1346 PIPPSVTOEBPPVPLAPLPAVDGQEBPPIPSAPAIATVOKSGSSTPALAGVLPP- 1402
OY 902 QGEKVEGLQALYPMKAKDNHLENFK-----SDVITVLEQQDMWFGEVQOK 951
Db 1403 --PPLPQOASTSEPTIAHDVNTNGAEKNGAYGSDSDVVLSPES-----VETDEEB 1455
OY 952 GWPEKSYVKLISGP 965
Db 1456 GAQPVSTAGIPISIP 1469

RESULT 10
YBET YEAST STANDARD; PRT; 1381 AA.
ID YBET YEAST
AC P34216;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 150.8 kDa protein in SEC17-QCRI intergenic region.
GN YBL047C OR YBL0520 OR YBL0501.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-961 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95176707; PubMed=7871888;
RA de Wergifosse P., Jacques B., Jomiaux J.-L., Purnelle B., Skala J.,
RA Goffeau A.;
RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast
RT chromosome II reveals homologues to bacterial proline synthetase and
RT murine alpha-adaptin, as well as a new permease and a DNA-binding
RT protein."
RL Yeast 10:1489-1496(1994).
RN [2]
RP SEQUENCE OF 579-1381 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94205266; PubMed=8154187;
RA Scherens B., el Bakhouy M., Vierendeels F., Dubois B., Messenguy F.;
RT "Sequencing and functional analysis of a 32,560 bp segment on the
RT left arm of yeast chromosome II. Identification of 26 open reading
RT frames, including the KIP1 and SEC17 genes."
RL Yeast 9:1355-1371(1993).
CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
CC -1- SIMILARITY: Contains 3 EH domains.
CC -1- SIMILARITY: Contains 1 UBA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1db-eb.ch/announce/
CC or send an email to license@1db-eb.ch).
CC -----
DR EMBL: Z35808; CA84867.1; -.
DR EMBL: X78214; CA85048.1; -.
DR EMBL: Z23261; CA880797.1; -.
DR PIR: S45781; S45781.
DR SCOP: S0000143; EDB1.
DR GO: GO:0005857; C:actin patch (sensu Saccharomycetes); IDA.
DR GO: GO:0005935; C:bud neck; IDA.
DR GO: GO:0005934; C:bud tip; IDA.
DR GO: GO:0006897; P:endocytosis; IMP.
DR InterPro: IPR000261; EPS15_homology.
DR InterPro: IPR000449; UBA_domain.
DR Pfam: PF00036; ehand; 2.

```

DR Pfam; PF00627; UBA; 1.
 DR SMART; SM00054; Efh; 2.
 DR SMART; SM00054; Efh; 3.
 DR SMART; SM00165; UBA; 1.
 DR PROSITE; PS00031; Efh; 3.
 DR PROSITE; PS00030; UBA; 1.
 DR Hypothetical protein; Repeat.
 FT DOMAIN 14 113 EH 1.
 FT DOMAIN 135 227 EH 2.
 FT DOMAIN 277 366 EH 3.
 FT DOMAIN 1338 1380 UBA.
 SQ SEQUENCE 1381 AA; 150783 MM; 626FD261DCBA7D99 CRC64;

Query Match 6.4%; Score 403; DB 1; Length 1381;
 Best Local Similarity 20.0%; Pred. No. 1.1e-08;
 Matches 266; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

3 QEPTEP-GGS-----LDVMTATVERAKHDQQLSLKPIAGFITGDARNFPSGHP 54
 109 QNPAPWQSGSATGNTNTNDI PALSSNDIAKFSQLPRTAKGQTVAGDKAKOIFLKARLP 168
 55 QGVLAQIYALAMNDGRMDQVEFSIAMKLIKLGQYQSTLPVPMKQAPVALISSAPA 114
 169 NQTLGEMALCDRDASGVLDKSEFFIMAWYLIQCMSHHPSMTTPPAVLPTQ----- 219
 115 FGIGGIASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPPAAVPLANGAPPVIOPLPAFAH 174
 220 -----LMDST-----RLEPVVNVQNPRTTPTLSANST 245
 175 PATWPKSSPSRSGSQSLNTKLOKQASFDVASAPPAEMAVPQSSRLKRYOLFNSHDK 234
 246 GVSLSLRHSITRLSLGAFSA-----ASDMSLFFKQKQDPAIFPSLDK 290
 235 TMSGHLTGQATITLMQSSLPQAOLASINLSDIDQKLTAEFTLANHLDIVANSGP 294
 291 QHAGSLSSAVLVPFLLSSRLNQTETATIMDLADIHNNAEFTKLEFAMFLQKNAGYE 350
 295 LPPVLPE-----YIPPSFRVRSGSGMSVSSSS----- 324
 351 LBDVLPNELLSGALGLYPNPLPQOQSAPOIALIPRASKPSLQDMPHQVSAVANTOPT 410
 325 VDQRLPE-----EPS-----SEDEQOP 341
 411 VPQVLPQNNGSLNDLALNPSFSSPPTKQTVVQNTNNSFSYDNNGCATLQOQOP 470
 342 EKKLPVTEFDKRENFERSGVLEKRRQALLPQKREERL----- 382
 471 QQPPPLTHSSSGLKFTPTS-----NFGQSI1KEEPPEQQLRESSDPTFSAOPPPVKHAS 526
 383 ----- 382
 527 SPVKRTASTTLPOVPNFSVSPMAGATSAATGAAGAAVGAALGASAFSSSSNNAFKN 586
 383 -----AQLERAQERKERERQEOEAKRQLEKOLEKORLELEROBEERKEIR 432
 587 QDLFADGEASQLSNATTEMALNSQVNSLSQASITNDKSRATBELKRVTEMKNSIOI 646
 433 REAAKRELERQOLEWERNRQELLNQRNKEQEGTVLAKRKTLEFELBALNDKKQOLE 492
 647 KLANLRSTHDQVQKQTEQLEAQL- QVANKENS-----TLAQQLAVEANVHAHE 694
 493 GKLQDIRGLATQROGIBSTNSRELRIMEI-----THLOOQLOESQOMGRILPEQQLIS 548
 695 SKLNE-----LTTDLQSSQTKNALKQKITNLNLSMTASLOSQNEKQO----- 737
 549 DQLKQVQONSILHARDSILLTKRALAEAKELARQOLEQLEVERETSKLOEIDVFNNOLKE 608
 738 -QVKQ-----ERSWVDVNSKQLELNQTVANLQKIDIG-----EKISVYLTKQOE 783
 609 LREIHSKQOLQKORSLEAARLKQKEOE--RKSLET--EKQKEDQARVQOERDKOMLEHV 663
 784 LND-----YQKTVBEQHOALQAKYQDLSNKTDLTDREKQOLEERRRQIEQEOENLVHQHV 837

QY 664 QOEOBRPRKPHBEDRLKREDSVRKKEABERAK-----PEMOKOSRLFPHQOE-- 713
 DB 838 SKLQF-----MFDLSQKASFEKADQELKERNIEYANNVRELSERQNL-AMGOLPED 890
 QY 714 -----AKLATQAPWST--TEKGPL--TIS-----AOESVYVYRYALYPRESSHDEI 757
 DB 891 AKDIIAKSASNTDTTTKAITSRGVNHEDTVSAFVETTVENSUNLVNVRVADDEKERTIS 950
 QY 758 TIQPGDI-VMVDESQTEGPMGLGELKGGTKGTFPANYAEKIP-----NEVPPT-APKV 809
 DB 951 DVFEDRDVPTLGSQSSSENANTNNGTOSGNETANP-NLTETLSDRFQDGLNEVGIPPSQSL 1009
 QY 810 TDLTAPAPKALRETPAPLPVTSSSEPTTPNNWADFSTWSSSNEKETNMDTWMAQ 869
 DB 1010 TSSVANNAAPQ-SVRD-DVELPTELEBDTI-WNTANRDT-----GNLSHIPEWEXTPTAT 1062
 QY 870 PSLTPSAGQLQRQSAFTPATATGSSPSVLG--QGEKVEGLQAO--ALYPWRAKQD- 923
 DB 1063 ASTDVLS-----NETTEVIDGSTTKRANGNEDGESVSSIQESPKLSAQ-KAKTIVE 1114
 QY 924 -----HLNFKSDVITYLEQODMMWFGEVQCKMPKSYVKLISGP----- 965
 DB 1115 EPPPIQELHIDSDSSDDDE--FEDTRE--ISATVXTLQTPYNAOPTSLSEIH 1166
 QY 966 ---VRKSTSIDGPTESPASILKRVASPAKPAIP--GEEFIMYTESSEQGLDITFQOQD 1020
 DB 1167 TEQVTKYPAAGTSPSHNBSGSKASTNLTLPKXDEPDEFAGL-----EQAAVEBDNG- 1219
 QY 1021 VIVVTKQGDMMWTGTVGKSGVFPANNVYALKDSE-----GSGTAGKTSLGKPEI 1071
 DB 1220 -----ABSESEFENVANANGSMEQFETIDHOLDDELQNNATGTLTSSNFTI-PKQV 1272
 QY 1072 AQVIASVYATGPEQULTPAGLILIRKKNPGSGWEGELQARKKQOIGFPPANYVKLSLP 1131
 DB 1273 QO-----QSTSDPAQV-----SNDEWD-EIIFA-----GF----- 1295
 QY 1133 GTSKITPTEL--PKTAVQAVQVIGWYDTQOND--ELARSKQIIN 1176
 DB 1296 GNSKAEPTKVATPSTIPQP-----IPLKNDPIVDASLSKQPIVN 1334

RESULT 11
 TRHY_RABIT
 ID TRHY_RABIT STANDARD; PRT; 1407 AA.
 AC P37709;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Trichohyalin.
 GN THH.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_Taxid=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fletz M.J., Rogers G.E.;
 CC Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.
 CC -FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
 CC -IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
 CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCARFOLD PROTEIN, TOGETHER
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
 CC DIFFERENTIATION.
 CC -SUBUNIT: Homodimer (Probable).
 CC -TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
 CC -DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
 CC THE EPIDERMIS.

[illegible]

RA MEDLINE=93101679; PubMed=14655449;
RX Hayashi K, Yano H., Hashida T., Takeuchi R., Takeda O., Asada K.,
RA Takahashi E.-I., Kato I., Sobue K.;
RT "Genomic structure of the human caldesmon gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:12122-12126(1992).
CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
CC REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
CC NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
CC FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
CC INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
CC TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.
CC THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
CC POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
CC MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
CC ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPING.
CC -1- SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
CC SKELETAL FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=1; Synonyms=H-CAD;
CC IsoId=Q05682-1; Sequence=Displayed;
CC Name=2; Synonyms=W1-38 L-CAD I;
CC IsoId=Q05682-2; Sequence=VSP_004155;
CC Name=3; Synonyms=HEILA L-CAD II;
CC IsoId=Q05682-3; Sequence=VSP_004154, VSP_004155;
CC Name=4; Synonyms=W1-38 L-CAD II, 1-CAD;
CC IsoId=Q05682-4; Sequence=VSP_004156;
CC Name=5; Synonyms=HEILA L-CAD II;
CC IsoId=Q05682-5; Sequence=VSP_004154, VSP_004156;
CC -1- TISSUE SPECIFICITY: High-molecular-weight caldesmon (isoform 1) is
CC predominantly expressed in smooth muscles, whereas low-molecular-
CC weight caldesmon (isoforms 2, 3, 4 and 5) are widely distributed
CC in non-muscle tissues and cells. Not expressed in skeletal muscle
CC or heart.
CC -1- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
CC BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
CC BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
CC HELICAL REGION IN THE SMOOTH-MUSCLE FORM.
CC -1- PPM: IN NON-MUSCLE CELLS, PHOSPHORYLATION BY CDC2 DURING MITOSIS
CC CAUSES CALDESMON TO DISSOCIATE FROM MICROFILAMENTS.
CC PHOSPHORYLATION REDUCES CALDESMON BINDING TO ACTIN, MYOSIN, AND
CC CALMODULIN AS WELL AS ITS INHIBITION OF ACTOMYOSIN ATPASE
CC ACTIVITY. PHOSPHORYLATION ALSO OCCURS IN BOTH QUIESCENT AND
CC DIVIDING SMOOTH MUSCLE CELLS WITH SIMILAR EFFECTS ON THE
CC INTERACTION WITH ACTIN AND CALMODULIN AND ON MICROFILAMENTS
CC REORGANIZATION (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CALDESMON FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64110; AAA5636.1; -
DR EMBL; M83216; AAA58420.1; -
DR EMBL; M83216; AAA58419.1; -
DR EMBL; D90452; BAA14418.1; -
DR EMBL; D90453; BAA14419.1; -
DR PIR; JH0628; JH0628.
DR Genew; HGNC:1441; CALDL1.
DR GK; Q05682; -
DR MIM; 114213; -
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0003779; F:actin binding activity; TAS.
DR GO; GO:0005516; F:calmodulin binding activity; TAS.
DR GO; GO:0005523; F:tropomyosin binding activity; TAS.
DR InterPro; IPR006017; Caldesmon_LSP.
DR InterPro; IPR006018; Caldesmon_LSP.
DR Pfam; PF02029; Caldesmon_1.
DR PRINTS; PRO1076; CALDESMON.

KM	Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
FT	Repeat; Alternative splicing.
FT	DOMAIN 319 375
FT	REPEAT 319 332
FT	REPEAT 333 346
FT	REPEAT 347 360
FT	DOMAIN 26 207
FT	DOMAIN 564 621
FT	DOMAIN 664 674
FT	DOMAIN 653 686
FT	DOMAIN 716 722
FT	DOMAIN 768 793
FT	DOMAIN 39 46
FT	DOMAIN 81 86
FT	DOMAIN 189 196
FT	DOMAIN 376 379
FT	DOMAIN 540 543
FT	DOMAIN 580 583
FT	DOMAIN 597 600
FT	MOD_RES 724 724
FT	MOD_RES 730 730
FT	MOD_RES 753 753
FT	MOD_RES 759 759
FT	MOD_RES 789 789
FT	VARSPLIC 1 24
FT	VARSPLIC 208 436
FT	VARSPLIC 208 462
FT	CONFLICT 530 530
SQ	SEQUENCE 793 AA; 93250 MW; 2A0DC63D16DD6B5F CRC64;
Query Match	5.3%; Score 330; DB 1; Length 793;
Best Local Similarity	23.5%; Pred. No. 3.9e-06;
Matches 150; Conservative 108; Mismatches 209; Indels 170; Gaps 25;	
OY	318 SVISSSSVDQLPEPPSEDEQQP-----EKKLVPTEDKKNENFERGVSYLEKRQ 369
DB	223 TVVMSLNKGGLISSEEPKEEREGSDIISHHEK--MEEDKRAEAERARLAEERRER 279
OY	370 ALLEQQRK-----EQRLAQLEAAEQEKREHQEQBAKQULENQLEQRLELR 420
DB	280 IKAQDDKKILDERARIAEERKAAQAQEERREAEERHMERREEKRAAE-EQRIKEEKRA 338
OY	421 QREERREKEIERREAARELERQORQLWERNRRDELNNKNEQGTVLKARR-KTLEF 479
DB	339 AEEGRRIKEEKRAAE----ERQRIKEEKRAAEERGAIAEERKAKVREQNKQOLE 394
OY	480 ELEALNDKQQLQESKLQDIRCLATOROEIES-----TKSRRLAIETTHLQOOLQES 533
DB	395 KKRAMOETIKIGEVYEOKIEGWNEKCAQBDKLTAVLKQGGEKGTQYAKKEKIQED 454
OY	534 QQMLGRLLPEKQLISDLQKVQONSLSHRDISLTTLKALEKELARQOLRQOLOVEHETR 593
DB	455 KPTE-----KKEEIKTD-----KTKQKEPE-----EVKSPMDKKGFTE 490
OY	594 SKLOEIVFNQNLKELREIHSK-----QOLQOR-----SLE 625
DB	491 VKSQNGEFMTHTKLGHNTENTSFRPGRASVDTKABEGA.PQYBAGRGLELARRRGETSEE 550
OY	626 AARKKQKEQERKSILE--KQKEDAQRVQDRDKOMLEHYVQEQBPAPRPKPHBEDRLKRE 683
DB	551 FEKKIQKQOO-AALIELELKKKEEREKRVLEEEOQ--RRQOEADRLKREEERKRLKEE 607

```

Oy 684 DSVRKEAEERAKPEMOD-----KQSLRPHPHQEPKATLQAPWSTTEKGLTISAGES 737
Db 608 IERRAEAEAKRQKPEDEGLSDDKKPFKCTPKGSLKIEERAE-----LINKSVOKS 660
Oy 738 --VVVVVYRAlY-PFESRSHDEITTPQDIWVDESGTGEPRQGLGELKGTGWPANYA 794
Db 661 SGVSTHQALVSKIDSLREQYTS-----AIEGTRKAKP----- 694
Oy 795 EKIPENVPPTAKPYTDL-----TSAPAP-----KIAL----- 822
Db 695 TKPAAASDLPVPAEGFRNKSMWKEGKNVSPFAAGTPKETAGLKVGVSRINEMWTXTYP 754
Oy 823 --RETPAPLPVTSSEPTTPNNMADFSSTWPSSEK 857
Db 755 DGNKSPAPKP-----SDLRPGDVSSKRNLMKQSVDK 786

RESULT 15
MAK4_HUMAN
ID MAK4_HUMAN STANDARD; PRT: 1239 AA.
AC O95819; OT5172; Q9NST7;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Mitogen-activated protein kinase kinase kinase 4 (EC 2.7.1.37)
DE (MAPK/ERK kinase kinase 4) (MEKKK 4)
DE (HRK/GCK-like kinase HGK) (NCK interacting kinase).
GN MAPK4 OR HGK OR NIK OR KIA0687.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 4 AND 5), FUNCTION, AND TISSUE
RP SPECIFICITY.
RC TISSUE=Macrophage;
RX MEDLINE=99107863; PubMed=9890973;
RA Yao Z., Zhou G., Wang X.S., Brown A., Diener K., Gan H., Tan T.-H.;
RT "A novel human STE20-related protein kinase, HGK, that specifically
RT activates the c-Jun N-terminal kinase signaling pathway.";
RL J. Biol. Chem. 274:2118-2125(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glycoblactoma;
RX PubMed=12612079;
RA Wright J.H., Wang X., Manning G., Lanvère B.J., Le P., Zhu S.,
RA Khatri D., Flanagan P.M., Buckley S.D., Whyte D.B., Howlett A.R.,
RA Bischoff J.R., Lipson K.E., Jallat B.;
RT "The STE20 kinase HGK is broadly expressed in human tumor cells and
RT can modulate cellular transformation, invasion, and adhesion.";
RL Mol. Cell. Biol. 23:2068-2082(2003).
RN [3]
RP SEQUENCE OF 362-1239 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ichikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kozumi H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [4]
RP SEQUENCE OF 262-1239 FROM N.A. (ISOFORM 3).
RC TISSUE=Testis;
RX Duecherhoef A., Lauber J., Mewes H.-W., Gaassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 700-1239 FROM N.A.
RC TISSUE=Brain;
RX Saito T., Seki N., Hori T.;
RT "Isolation, expression profile and chromosome assignment of a novel
RT serine/threonine kinase gene.";
RL Submitted (Apr-1998) to the EMBL/Genbank/DBJ databases.

```

```

CC -1- FUNCTION: Serine/threonine kinase that may play a role in the
CC response to environmental stress and cytokines such as TNF-alpha.
CC Appears to act upstream of the c-Jun N-terminal pathway.
CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphorylated protein.
CC -1- COFACTOR: Magnesium.
CC -1- SUBUNIT: Interacts with the SH3 domain of the adapter proteins
CC NCK (by similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=1; Synonym= tumor-associated;
CC IsoId=O95819-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95819-2; Sequence=VSP_007054, VSP_007057;
CC Name=3;
CC IsoId=O95819-3; Sequence=VSP_007056, VSP_007057;
CC Name=4; Synonym=HGK-S;
CC IsoId=O95819-4; Sequence=VSP_007054, VSP_007055, VSP_007057,
CC VSP_007058;
CC Name=5; Synonym=HGK-L;
CC IsoId=O95819-5; Sequence=VSP_007054, VSP_007055, VSP_007056,
CC VSP_007057, VSP_007058;
CC -1- TISSUE SPECIFICITY: Appears to be ubiquitous, expressed in all
CC tissue types examined. Isoform 5 appears to be more abundant in
CC the brain, isoform 4 is predominant in the liver, skeletal muscle
CC and placenta.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -1- SIMILARITY: Contains 1 CNH domain.
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
CC frameshift in position 1151.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.jdb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF096300; AAD16137.1; -
CC EMBL; AY212247; AAO32626.1; -
CC EMBL; AB014587; BAA31662.1; -
CC EMBL; AL137755; CAB70907.1; ALT_FRAME.
CC EMBL; AB013385; BAA33714.1; -
CC Genew; HGNC:6866; MAPK4.
CC MIM; 604666; -
CC GO; GO:0005524; F:ATP binding; IDA.
CC GO; GO:0004674; F:protein serine/threonine kinase; IDA.
CC GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
CC GO; GO:0007243; P:protein kinase cascade; IDA.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR001180; Citron.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_kinase.
CC Pfam; PF00780; CNH; 1.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00036; CNH; 1.
CC SMART; SM00220; S_TKC; 1.
CC SMART; SM00219; TYRK; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC ATP-binding; Serine/threonine-protein kinase; Transferase;
CC KW Alternative splicing.
CC DOMAIN 25 289
CC FT NP BIND 844 1143 CNH.
CC FT BINDING 31 39 ATP (BY SIMILARITY).
CC FT ACT_SITE 54 54 ATP (BY SIMILARITY).
CC FT VARSPLIC 153 153 BY SIMILARITY.
CC FT 495 525 Missing (in isoform 2, isoform 4 and
CC /FTId=VSP_007054.

```

```

FT VARSPLIC 569 622 Missing (in isoform 4 and isoform
FT FT 5).
FT VARSPLIC 623 623 /FTid=VSP 007055.
FT VARSPLIC 623 623 V -> VQSHSLAKNVSPVSRSHSPSPKFAHHHRS
FT VARSPLIC 623 623 QDCPPSRSEVLQSDSDSKSEADPTQKXMSDSDVEVPR
FT VARSPLIC 623 623 V (in isoform 3 and isoform 5).
FT VARSPLIC 623 623 /FTid=VSP 007056.
FT VARSPLIC 623 623 A -> AGEV (in isoform 2, isoform 3,
FT VARSPLIC 623 623 isoform 4 and isoform 5).
FT VARSPLIC 623 623 /FTid=VSP 007057.
FT VARSPLIC 623 623 H -> HVRKNPHSM (in isoform 4 and isoform
FT VARSPLIC 623 623 5).
FT VARSPLIC 623 623 /FTid=VSP 007058.
FT VARSPLIC 623 623 S -> SS (IN REF. 3 AND 4).
FT VARSPLIC 623 623 O -> R (IN REF. 2).
FT VARSPLIC 623 623 O -> R (IN REF. 2).
SQ SEQUENCE 1239 AA; 142100 MW; 8FB8E2F9ABEC757 CRC64;

Query Match
Best Local Similarity 22.1%; Score 330; DB 1; Length 1239;
Matches 193; Conservative 116; Mismatches 222; Indels 342; Gaps 39;

QY 331 EEPSSDEQPEKKLPYTFEDKKRENFERSVLEKROALLQCKEORLAQLERAQ 390
DB 330 EEEVPEQEGEPSSIVNVPESTLRDPLRLOENKERSALRQQLQEQOL----- 381
QY 391 EKKEREROEAKROL--ELEKOLEKORLEROEERREKETERREAKRELEOROLEM 448
DB 382 -----REOEYKROLAROKRIEQCKORRLEEQORRE--REARRQOEBOORREQ 432
QY 449 EENRROELNORNEQEGTVLVKARKRTLEFELEALNDKHOLEGLODIRCLATOROE 508
DB 433 EKKRRLLEBERRKEE-----ERRAEEKRYERQEIYIRQLBEEGRH 478
QY 509 IESTKSRRLRLAETHLOOQLOESQOMGLRPEKQILSDOLKOVQNSLHRDSLTLTK 568
DB 479 LEV-----LOOQLQEQAML----- 493
QY 569 RALEAKELARQOLREQLDVERETRSKLQELDVFNQQLKELREIHSKOLOKORSLEAAR 628
DB 494 --LECR-----WRMEBHROAERLQRO----- 513
QY 629 LKQKQERKSLLEKOKEDAKORRVOERDKOMLEHVQOEBOOPRPRKPEBEDRLKREDSYRK 688
DB 514 LQOEQAYLLSLQHDHRRPHQHSQOPP-----PQERKSPFHAPEPKAHYEPADRA 566
QY 689 KEAER-----AKPEMDKO-----SRLFHPH-QEPKATATQAP 721
DB 567 REVEDRFRKTNHSSPEAQSKQGRVLEPPVPSRSESFSNGNSESVHAPALQRPAP--PQVP 624
QY 722 WSTTEKGPL-----TISAGESVKVYVYRALYFPESHDEITIQPDIVM 766
DB 623 VRTTSSPVLARRDSEPLQSGQOQNSQAGRNSTSTIEPRLLW--BRVEKLVPRPGSSS 680
QY 767 VDESQGTG-EPGWLGGELKG-----KTGWFPPANYA--KIPEN--EVTPEAKPV 809
DB 681 SSSSNSGSGQPGHPSGSGSGERFRVRSKSEBSPSQRLNNAVKKPEDKXEVFRPLKP- 739
QY 810 TULTSAPAKLALRETPALPYT---SSEPSTPEN-----NMADPSSSTWP----- 851
DB 740 ADLITLAKELRAVEDVRPHKYVDYSSSESGTDEEDDVEQEGADESTSGPEDTRAA 799
QY 852 SSSN-EKPETDN-----WDTMAOAPSLVPSAGOL-----RQSAFTP- 888
DB 800 SSLNLSNGETESVKTYIHDVDESEBAMTPSKEGTLIVROTOSASSTLQKHKSSSEFTPF 859
QY 889 -----ATATGSSPSPVLAGOEKVEGLQAQALYPRAKDNHLNFKNSDVITVLEQOD 940
DB 860 IDPRLQLGISPSGTTVTSV--GFSCDGRPEAL-----RQD 894
QY 941 MMHFGVQOGQKMGFKSYKLSGPYKSTSIDTGPTES-PAS-----LKRVASPAA 991
DB 895 -----PTRKGSVVVNPNTNRPQSDTPEIRKYYKRFENSEIL 930

```

```

QY 992 KPAIPGEEFI-----AMTYESSFQGD-----TFOQGDV-----IVT---KKG-- 1029
DB 931 CALMGVNLVGTESGMLNDRSGGKRYPLINRRFQOMDVLBGLNVLVTISGKDKLR 990
QY 1030 ----DW-----WTGTVGDSKG 1041
DB 991 VYLSWLNKILANDPEVEKKQGT-TVGDLLEG 1022

```

Search completed: December 4, 2003, 15:19:11
Job time : 22 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:11:28 / Search time 36 Seconds
(without alignments)
8694.939 Million cell updates/sec

Title: US-09-674-237A-3
Perfect score: 6269
Sequence: 1 MAQPPFGSGSIDVNAITVE.....VGLFSPSNVYKLTITMDPSSQ 1213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SP archaea:*
 - 2: SP bacteria:*
 - 3: SP fungi:*
 - 4: SP human:*
 - 5: SP invertebrate:*
 - 6: SP mammal:*
 - 7: SP mhc:*
 - 8: SP organelle:*
 - 9: SP phage:*
 - 10: SP plant:*
 - 11: SP rodent:*
 - 12: SP virus:*
 - 13: SP vertebrate:*
 - 14: SP unclassified:*
 - 15: SP viirus:*
 - 16: SP bacteriaph:*
 - 17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the target being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4264.5	68.0	1220	13 Q8UFT5	Q8UFT5 brachydanio
2	4264.5	68.0	1721	13 Q8UFT4	Q8UFT4 brachydanio
3	1738.5	27.7	1014	5 Q8INU2	Q8INU2 drosophila
4	1735	27.7	1011	5 Q61639	Q61639 drosophila
5	1702	27.1	1097	5 Q9VIF7	Q9VIF7 drosophila
6	1698.5	27.1	1094	5 Q61618	Q61618 drosophila
7	1684	26.9	345	4 Q81WH9	Q81WH9 homo sapien
8	1530	24.4	611	4 Q81W15	Q81W15 homo sapien
9	1333	21.3	1097	5 Q9U2T9	Q9U2T9 caenorhabdi
10	1314	21.0	755	11 Q8C4B5	Q8C4B5 mus musculu
11	1172	18.7	440	11 Q8C9C3	Q8C9C3 mus musculu
12	822.5	13.1	164	11 Q8C143	Q8C143 mus musculu
13	812.5	13.0	158	11 Q8C6U5	Q8C6U5 mus musculu
14	794.5	12.7	662	5 Q8T068	Q8T068 drosophila
15	704	11.2	276	11 Q8CD59	Q8CD59 mus musculu
16	536.5	8.6	864	4 Q9UBC2	Q9UBC2 homo sapien

17	531	8.5	1253	5 Q9W111	Q9W111 drosophila
18	528	8.4	1253	5 Q8W061	Q8W061 drosophila
19	527	8.4	907	11 Q60902	Q60902 mus musculu
20	526.5	8.4	1232	5 Q8WMD2	Q8WMD2 drosophila
21	510.5	8.1	819	11 Q8CB70	Q8CB70 mus musculu
22	506.5	8.1	751	5 Q23418	Q23418 caenorhabdi
23	506.5	8.1	796	5 Q9B1F4	Q9B1F4 caenorhabdi
24	504.5	8.0	1106	5 Q8WMD3	Q8WMD3 drosophila
25	496.5	7.9	599	11 Q8CB60	Q8CB60 mus musculu
26	493.5	7.9	599	11 Q91WH8	Q91WH8 mus musculu
27	481	7.7	681	5 Q81FX4	Q81FX4 caenorhabdi
28	478	7.6	108	11 Q8CJ62	Q8CJ62 mus musculu
29	434.5	6.9	1116	5 Q9HGL2	Q9HGL2 schizosacch
30	405	6.5	944	5 Q18137	Q18137 caenorhabdi
31	405	6.5	1148	5 Q95Z77	Q95Z77 caenorhabdi
32	405	6.5	1439	5 Q8WQC7	Q8WQC7 caenorhabdi
33	405	6.5	1480	5 Q18138	Q18138 caenorhabdi
34	402.5	6.4	1003	12 Q91LX9	Q91LX9 kaposi's sa
35	397	6.3	1089	12 Q40947	Q40947 kaposi's sa
36	387.5	6.2	826	5 Q9VPS3	Q9VPS3 drosophila
37	386	6.2	1036	12 Q9DUM3	Q9DUM3 kaposi's sa
38	386	6.2	2376	5 Q9V5J0	Q9V5J0 drosophila
39	386	6.2	2376	5 Q966V1	Q966V1 drosophila
40	383.5	6.1	1909	5 Q25893	Q25893 plasmodium
41	382	6.1	1162	12 Q98148	Q98148 kaposi's sa
42	377.5	6.0	1129	12 Q9OR71	Q9OR71 kaposi's sa
43	375	6.0	1151	4 Q9BX65	Q9BX65 homo sapien
44	373	5.9	1596	5 Q81J44	Q81J44 plasmodium
45	367	5.9	976	12 Q9DUN0	Q9DUN0 kaposi's sa

ALIGNMENTS

RESULT 1

Q8UFT5 PRELIMINARY: PRT: 1220 AA.

AC Q8UFT5

DT 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE SI:d2173A8.1.2 (Novel protein similar to interseclin (SH3 domain protein, ITSN1)).

GN SI:D2173A8.1.

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RA Pandian R.;

RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 5 SH3 DOMAINS.

DR EMBL; AL606751; CAD43427.1; -

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000261; EFS15_homology.

DR InterPro; IPR00108; Neu_cyt_fact_2.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00036; efhand; 2.

DR	PRINTS; PR00499; P67PHOX.
DR	PRINTS; PR00452; SH3DOMAIN.
DR	PRODOM; PD000066; SH3; 5.
DR	SMART; SM00054; EFh; 2.
DR	SMART; SM00027; SH3; 2.
DR	SMART; SM00326; SH3; 5.
DR	PROSITE; PS00018; EF_HAND; 2.
DR	PROSITE; PS00031; EFT_2.
DR	PROSITE; PS00002; SH3; 5.
KW	SH3 domain.
SQ	SEQUENCE 1220 AA; 136351 MW; 1AB9134E2A4BA515 CRC64;
Query Match	68.0%; Score 4264.5; DB 13; Length 1220;

Best Local Similarity 68.3%; Pred. No. 1.5e-206;
Matches 869; Conservative 108; Mismatches 184; Indels 111; Gaps 21;

QY	MAOPEPTFGGSLDVMATIVERARHDOOFLTKPL-AGFIIGDOARNFPFQSG,POPLTA	59
Db	1 MAOPEPTFGGSDTWISVDERAKHDOOFSHLTPPAGFITGDQARNFLOSGLPAPILTA	60
QY	Q1MALADMNNDGRDYOEFISIAMLKIKLQOQYOLPSTLPVPMKOQPVALISABAFIGG	119
Db	Q1MALADMNNDGRDYMHEFISIAMLKIKLQGHPLPSPMSMQOPL-----	108
QY	120 IASNPRLTVA PVPMGSI PVPMGSPPLVSVSPRAAVPPLANGAPVIOPLPAHPAATW	179
Db	109 --SLTP-----PTBFG--VPGMGVSPPLGASVTP--PVPSLAGAAMIOPLSGFHPGVLV	159
QY	180 PKSSFSRSGGSOQLNTLQAKOSFDVVS---APPAAMVAVPOSSRLKYROLFNSHDKTM	236
Db	160 NKTSGFNRS-----SITLQOGQSVAPSAAPAPPADMAVOSGRKYROLFNSHDKTM	213
QY	237 SGHLTGPARTILMQSSLPOAQLASIVNLSDIDODGKLTAEFFILAMHLIDVANGOPLP	296
Db	214 SGYLTGPARTILMQSSLPOAQLATIVNLSDIDODGKLTAEFFILAMHLIDMANGSLPLP	273
QY	297 PVLPPEYIPBEPFRVRSRSGMSVYSSSVDOQLPEBPSBEOBPBCKLPTPFEDKKRN	356
Db	274 PLPLPDLPLPTFRNRSSGVSIVTSMHSTD,RSOEPP--EBBEKAEKKLPTPFDKREN	332
QY	357 PERSVLEKROALBEOGRKEOBLAOLEBAOERKREBOEOBARKQLELEKOLEKOR	416
Db	333 PERGNLEKROALLBEOGRKEOBLALEBOERKREKRELBEOERRQOELDQLEKOR	392
QY	417 ELEBOREBERKEIERRBAARELEBOROLEMERNRQELLNQRNKEOGTVLKARKKT	476
Db	393 ELEROREBERKEIERREAAKRELEBORQLEMERTRQOELLTQRNREGENIVLTKARKKT	452
QY	477 LEFELBALNDKKHOLEGKLODIRCLATOBORBIESTNKSRELRIAEITHLLOOQOESQOM	536
Db	453 LEFELBALNDKKSQLEGLDORISRLS,ORHEIETNTKRELRIAEITSLQOOQOESQOM	512
QY	537 LGRILPEKQILSDOLKOYQONS,LRHDS,ILTLTKRALAEALKEARQOLREOLDVEBERSTRK	596
Db	513 LSRLLPBDQCLNDQKOYQONS,LRHDS,INS,ORAIEMKESTRKQOLKEJDAVEKETRSL	572
QY	597 QEIDVFNNOLEKREIHSKQOLQOKORSLEAARLQKOEGRKSLBEEKEDAOBRVDERD	656
Db	573 LEIDAFNTQLKELREIHNKQOKOKELE-ADITQPSHDKSIDBODSR-----	620
QY	657 KQMLEHVQOEOBP-----RPKKHPEEDRLKREDSVRKK--EAEERAKPEMOKOS	704
Db	621 ---LSGTTDGVSPAWRDGLKALPPPSQAMMSR---VREKKIVDARHGFDFCCCFM	674
QY	705 RLPHRPOPAKLATQAPMWSTKEGGLTISAOESVYVYRYALYPESHESHOETIOPGDI	764
Db	675 RVNVRNQGKSCVC-----VLDKLPVSGPNQEKVYVYRYAMYPPEAKSHBEITIHPEDI	728
QY	765 VMVDESQTBEGMWLGELKGTGMFPANYAEKIPENEVPYTPAKPVDTJLSAP-----	816
Db	729 VMVDESQTBEGMWLGELIKGTGMFPANYAEKIPSEVPLSLRASAASSAPKLGSHMS	788
QY	817 -----APKALARETPAL-PTVSSSESTTP-----NNMADFSTWSSSN-EKPE	859
Db	789 ASSSTTATPILPVSTEPSIASIPSSAPPTGPASSSSASSANMADFSTWSSNSA,VEKOD	848
QY	860 TDNMDDTW---AAOPELTVPSAGOLRORAFAPATATGSSPSPVJQOGKEVGLQAOALYP	916
Db	849 SDGMDAWTQTOPSLSVPSGGQIKORSAFTPATISSGSPBPVLQOGKEVGLQAOALYP	908
QY	917 WRAKDNHLNFKNSDVITVLEQODMWPFGEVQOGKMPKSYVKLISGPVKSSTSIDTP	976
Db	909 WRAKDNHLNFKNDVITVLEQODMWPFGEVQOGKMPKSYVKLISGPVKSMSIDSGS	968
QY	977 TESPASLKRVASPAKPA-----IPGEFIAMNTYSSBEGDILTFQOGDYIV	1022

```

Db      969  SDSPSPYKRPSPSLNKPTELGEGQNSNSNMYPKEVYAAITYAESNQGLTLFQGGDVT 1028
Qy      1024  VTKKDGMMWMTGYCWDKSGVFPSPNYVRLKDBEGSGTACKTSGSLGKKPPIAIVIA5YAAATGP 1083
Db      1029  VTKKEGMMWMTGYVSGKLGVPSPNYVRKPEBEGLSACKTSGSLGKKPPIAIVIAIYATYATGA 1088
Qy      1084  EQLTLAEGQILIRKKNKPGGMEGELQARGKKQIGWFPANYVRLSPGTSKIRPTE--TL 1141
Db      1089  EQLTLAEGQILIRKKNKPGGMBEELQARGKKQIGWFPANYVRLSPSTSKITPTDPPNP 1148
Qy      1142  PKTAVQPAVCQVIGMIDYTLQNDDELAFSGQIILNVLNKEDPDMWKEEVSQGVGLFPSSNY 1201
Db      1149  PKLPTPAVAVCOVIGMYDYTLQNDDELPEFGKQIILNVLSPREDPDMWKEELNGSVGLFPSNY 1208
Qy      1202  VKLTTMDMPSQO 1213
Db      1209  VKLTTDTDPSSQO 1220

```

RESULT 2

```

ID Q8BFT4 PRELIMINARY; PRT; 1721 AA.
AC Q8BFT4
DT 01-OCT-2002 (TREMBLrel, 22, Created)
DT 01-OCT-2002 (TREMBLrel, 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel, 23, Last annotation update)
DE SI:dz173A8.1.1 (Novel protein similar to intersectin (SH3 domain
DE protein, ITSN1)).
EN SI:dz173A8.1.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
CX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pardian R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 SH DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
DR EMBL; AL606751; CAD34428.1; -.
DR InterPro; IPR000008; C2.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS1_homology.
DR InterPro; IPR000108; Neu_cyt_fac2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhogEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00036; ehband; 2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhogEF; 1.
DR PRINTS; PR00459; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 5.
DR SMART; SM00239; C2; 1.
DR SMART; SM00054; EFh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhogEF; 1.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS50004; C2_DOMAIN_2; 1.
DR PROSITE; PS50010; DH_2; 1.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 5.
DR SH3 domain.
SQ SEQUENCE 1721 AA; 194115 MW; 2855789549DC07 CRC64;
Query Match 68.0%; Score 4264.5; DB 13; Length 1721;
Best Local Similarity 68.3%; Pred. No. 2.3e-206;

```

Matches 869; Conservative 108; Mismatches 184; Indels 111; Gaps 21;

```

QY 1 MAQPTFGSGSLDVAITVEERAKHDQQLSLKPI-AGFITGDAQANFFGSGLPQVLA 59
DB 1 MAQPTFGSGSDTWISYDERAKHDQQLSLTPPAGFITGDAQANFFGSGLPQVLA 60
QY 60 QIUALADNNNDGRMDQVEFSIAMKLIKLOQYOLPSTLPVWKOQPVAISSAPAGIG 119
DB 61 QIUALADNNNDGRMDHFEFSIAMKLIKLOQHPLPPLSPWKOQPI----- 108
QY 120 IASMPPLTAVAVPMKSIPIVWGMSPPLVSSVPRAVPLANGAPPIQOLPAPAPAAATW 179
DB 109 --SIP-----PTPFG-VPMGVSPLGAVTP-PPVPSLANGAPPAIQLPGLSGSHGAVL 159
QY 180 PKSSPFRSGPSQLNTKLOKXASFPVAS---APPAEMAVPQSSKTKROLFNSHDKM 236
DB 160 NKTPEFNRS-----SILQKQGSVAPAPAPAPPADMAVSSSLKTKROLFNSHDKM 213
QY 237 SGHLTGPOARTILMOSLSLPOAQLASIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGQPLP 296
DB 214 SGYLGPQARTILMOSLSLPOAQLATIMNLSIDIDQDKLTAEFFILAMHLIDVAMSGLP 273
QY 297 PVLPEYIPPSFRVAVSGSGMYSVSDQRLPEPSSSEDOQPEKLPVTFEDKREN 356
DB 274 PLPLPDLIPPTFRMRKSGSVSTSMHSTDLRSQEEP-EEEKEAKKLPVTFEDKREN 332
QY 357 PERGSVELKRRQALLQOKRQOERLAOLERAQERKEREQERQOERLELEKQER 416
DB 333 PERGNLEKRRQALLQOKRQOERLAALAEKQERKEREQERQOERLEKQER 392
QY 417 ELERQERERKEREEREAKELEEROROLEWERNRQELLNQNRKQEGTVLAKARKT 476
DB 393 ELERQERERKEREEREAKELEEROROLEWERTROQLNQRNEQENIVLAKARKT 452
QY 477 LEPELEALNDKXQLEGLQIDTRCRLATROEITESTNKREIRIATITLQOOLQESQOM 536
DB 453 LEPELEALNDKXQLEGLQIDTRSRISIRHETETTKREIRIATITLQOOLQESQOM 512
QY 537 LERLIPKQILSDQKQVQONSILHRPSLTLTKRALEAKELAQOALEQOLDEVERERSTL 596
DB 513 LERLIPKQILSDQKQVQONSILHRPSLTLTKRALEAKELAQOALEQOLDEVERERSTL 572
QY 597 QEIDVFNQKLELEIRHSKQOLQOKORSLBAARLKQKQERKSTLEKQEKEDAQORQERD 656
DB 573 LEIDAFNQLKELREIHNNQOKQOKQELE-ADITQSHRKISIDSQDSR----- 620
QY 657 KQMLEHVQOQEEP-----RPRKHEEDRLKREDSVAKK---EAEERAKEMQDKOS 704
DB 621 ---LSTGDDGVBPAMDGLGKAPPTPVGQAMMSR---VREKKIVDAERHGFQCCCFM 674
QY 705 RLPHHQEPAKLATQAPMSTTEKGLTISAQESVKVYVYRALYPPFSRSHDETTOPGDI 764
DB 675 RVNVRNQHSQVC-----VLNDIPVSGNQEKVKVYVYRAMPFARSHDETTIHPGDI 728
QY 765 VNVDESQTEPGMLGELKGTGMPFANYAEKIPENEVETPAKPVYDILTASP----- 816
DB 729 VNVDESQTEPGMLGELKGTGMPFANYAEKIPESVPLSLRASAASSAPRLGSHMS 788
QY 817 -----APKALRETAPPL-PVTSESEPTP-----NNWADFSSTWMPSSSN-EKPE 859
DB 789 AASSTTATPILPVSTEPASIAFPSSAPPTGPASSSSSASNMWADPSTTWMPSSNAVEKQD 848
QY 860 TQNWQDWT---AAOPSLTVPSAGQLRORSAFTPATATGSSPPLVGGGEYEGLOAALY 916
DB 849 SGGWMDWPTQPTQPSLTVPSGQIRORSAFTPATITLSSSSPVLGGGEYEGLOAALY 908
QY 917 MPRAKDNHILNFNSKSDVITVLEQODMMWFGVQOGKMPFKSYKLSGPRKSTSIDTP 976
DB 909 MPRAKDNHILNFNSKSDVITVLEQODMMWFGVQOGKMPFKSYKLSGPRKSTSIDTP 968
QY 977 TESPAKLKVASPAAPPA-----IPGEFIAMTYTESQGLTFFQGGDVIT 1023
DB 969 SDSPSPVKKPSPILNKPTLEGQSGNSNSMYSKEYVAMTYTESNEQGLTFFQGGDVIT 1028

```

QY 1024 VTKKGDMMWTGTVGDKSGVFPSPNVYRLKDSBSSGTAGTSGLSGKKPEIAQVIASAAATGP 1083

DB 1029 VTKKGDMMWTGTVGDKSGVFPSPNVYRLKDSBSSGTAGTSGLSGKKPEIAQVIASAAATGP 1088

QY 1084 EQLTLAPGQLIIRKKNPGQWEGELQARKKRQIGMPFANYKLLSPSTSKITPTE--L 1141

DB 1089 EQLTLAPGQLIIRKKNPGQWEGELQARKKRQIGMPFANYKLLSPSTSKITPTE--L 1148

QY 1142 PXTAVQAPVCQVIGWYDTAQNDELAFSGQITNVLKKEDPDMKKGVSQGVLPSPNY 1201

DB 1149 PXLPPNAPVCQVIGWYDTAQNDELAFSGQITNVLKKEDPDMKKGVSQGVLPSPNY 1208

QY 1202 VKLTTDMPDPSQ 1213

DB 1209 VKLTTDTPDPSQ 1220

RESULT 3

OSINUZ

ID OSINUZ PRELIMINARY; PRT; 1014 AA.

AC OSINUZ;

DT 01-MAR-2003 (Tremblrel. 23, Created)

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE CG1099-PB.

GN DAPI60.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_Taxid=7227;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe O., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,

RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertman B.P., Bhandari D., Bolshakov S.,

RA Borzova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davernport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostel A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegyam C.,

RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasbo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mikhlov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy D., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T.,

RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svitakae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.Y., Wassenaar D.A., Weinstein G.M., Weisenbach J.,

RA Williams S.M., Woodgett W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,

RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195 (2000).

[2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Batzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Fartan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
 RA Paciel P., Parag S., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pitman G.S., Puri V., Richards S., Scheeler F.,
 RA Stepieton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri U.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE003669; AAN11099.1; -;
 SQ SEQUENCE 1014 AA; 111890 MW; 633790A316865FB4 CRC64;

Query Match 27.7%; Score 1738.5; DB 5; Length 1014;

Best Local Similarity 33.5%; Pred. No. 1.7e-79;

Matches 414; Conservative 204; Mismatches 332; Indels 285; Gaps 31;

QY 11 SLDVAITVERAKIDQOPLSLKPIAGITGDQANFFPQGLPPVLAQIMALDMND 70
 Db 4 AYDAVAVPRERELKYOQFRLQPGAGVTGAQAGFLQSQPLILIGQIMALADTSD 63
 QY 71 GRMDVERSIAMKIKLKLOGLYLPSTLPVVKQOPVAISSAPAGIGIASMP---LT 127
 Db 64 GGMNINIESIACKLNLKRGMDVKVLPSSL-----LSSL-----TEDVFSMTFRGSTS 113
 QY 128 AVAPVPMGSIPIVVGSPPLVSSVPPAAVPPPLANGAPVIOPLPAFAHPAATWPKSSFSR 187
 Db 114 SLSPLD-----PLKGI-VPAVAVPVVAVPVA--VATVISP-PGVSVSPGPPPTSN--- 162
 QY 188 SGPGSLNTKLOKQSFVDAAPPAEWAANVPOSSSLKTRQLFNSHDKTMSGHLTPQART 247
 Db 163 --PGRHRTSISERAPSIIESVN---OGEWAVQAQKRTQYVNAADRRTSGYLTSQARG 217
 QY 248 ILMGSLQOALASTWNLSDIDQDKLTAEFFITLMHILIDVMSGQPLPPVLPPEYIPPS 307
 Db 218 VLVOGKLPQVTLAQITLSDIDGRNLNDEFILTMFLCEKMAAEKIPVTLPOEWPEPN 277
 QY 308 FRFVSGSGMSVYSSSVDOPLREPPSSD-----EQQPEKLP--VTFEDKK 353
 Db 278 LRKISRG---SVSGVAVSRGSPASHASVSSQSGVGVADADPTGLPGQTSFEDKR 333
 QY 334 RENFRGSGVELEKRRQALIEQKQOEKRLAQLEFRAEQEKERQOEKAKRLIEKQLE 413
 Db 334 KENYVGAELDRRRKIMEDQOKERERERERERERERERERERERERERERERERER 393
 QY 414 KORELERQEEER 473

Db 394 ROEIEWEKEQRKRELEAKRELEKORQOEWEQARIENNAQREERQEVHLKOKAH 453
 QY 474 RKLIEFELELNKQKQLEBKODIRCLATQOEISTKSRERLAEITHLQOQOES 533
 Db 454 NTOANLESTLNKIKELSORICDTRAGVNVVTVIDGMETQDTSSEWSQLARKIEQ 513
 QY 534 QOMGLRIPK-----OILSPOLQVOONSILHRDSTLTLKRLAKELARQOLR 582
 Db 514 NAKLUOTQORAKREAKSKASGAALGEMNAQOE-----LNAFAHKOILINQIK 563
 QY 583 EOLDEVERETRSKLOEIVENNOLKEIRE-----HSKQOLQORSLEAARLKOEKERK 637
 Db 564 DKVENISKETIESKEDINTNDVQMSSELKALSALITCEPLLYXEYDQRTSVLELKNRK 623
 QY 638 SLEEKKEQAKRVRQCRDKOMLEHVOEEOPRKRKEHEDRLKREDSVRKKEAKERAKP 697
 Db 624 NETSVSSAMPT-----GSSSAW-----EEGTVTYDYY-----AVASNDLSALAAP 664
 QY 698 EMODKQRLPHPHQEPKATQAAPWSTTEKGPLTISAQESVYVYRALYPFESRSHDEI 757
 Db 665 AVD-----LGGPAP-----EG-----PVKYAVVEFNARNAEI 693
 QY 758 TIOPGDIVMYDESGTGPGLGELKKGKGTWFPANVAEKIPENEVPTPAKPVDTLTSAPA 817
 Db 694 TFVPGDILLVPLEQNAEPGWLAGEINGHTGMFPESYVEKLEVGCV-----A 739
 QY 818 PKALRETPAPLPTSSSEPTTPPNMADFSSTWSSSEKPEPDNDMTMAAPSLTVPSA 877
 Db 740 PVAAVE---APVDAQVATVADTYN-----DNINT-----SSIPAA 771
 QY 878 GQLRQSAFTPATGTGSSPSPVLQCGEKVEGLQALYPRAKKNDLNFNKSQDITVLE 937
 Db 772 S-----ADLTPAG-----DV----- 781
 QY 938 QQDMWMEGVQGXGWPFPKSYVKLISGPVRKSTIDGTSPASLKRVASPAKPAIPG 997
 Db 782 ----- 781
 QY 998 EEFIAMTYESSEGDITFQOQGVIVYTKKDDGMWTTVGDKSGVPSNRYRLKDSBSG 1057
 Db 782 EYVIAAPYSEAEAGDLSFSAGENVWVYKKEGEMWTTISRTGMFPNSNVQKADVGTAS 841
 QY 1058 TAG-----KTSGLGKKPEIAQVIASVYATGPQLTLAPQOLILIRKKNPGKMGESLQ 1110
 Db 842 TAAEPVESLDQGRARAKSEIAQVIAYEATSTQLTLTGQILMIRKKTDSGMBESLQ 901
 QY 1111 ARGKRRQIGWFPANVYKLLSPG--TSKITPTLEPKTAVQPAVC--QVIGMDITYAQNDEL 1167
 Db 902 AKGRRRQIGWFPATYVAVLQGRNSGRNTPVSGSRIENTEQILDVKIALYFYKAQNDEL 961
 QY 1168 AFSKQIINVLNKEDPDWKGGEVSGQVGLPPSNV 1202
 Db 962 SFDKDDIISVIGRDEPEWMRGELNGLSLGPSNV 996

RESULT 4
 ID 061639 PRELIMINARY; PRT; 1011 AA.
 AC 061639;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Dytamin associated protein isoform DAPI60-2.
 GN DAPI60 OR CG1099.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydridae; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R; TISSUE=Head;
 RA Roos J., Kelly R.B.;

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kratke C., Kravitz S., Kulp D., Lai Z.,
RA Laebo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusekern D.R., Pacleb J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkova R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195 (2000).
CC 1. SIMILARITY: CONTRAINS 4 SH3 DOMAINS.
DR EMBL; AE003669; AAF53962.1; ..
DR HSSP; P29354; IGRI.
DR Flybase; FBgn0023388; Dap160.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_homology.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00036; ehfand; 2.
DR Pfam; PF00018; SH3; 4.
DR PRINTS; PR00452; SH3DOMAIN.
DR Prodom; PD000066; SH3; 4.
DR SMART; SM00054; Eph; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS00018; EF_HAND; 2.
DR PROSITE; PS50031; EH; 2.
DR PROSITE; PS50002; SH3; 4.
DR KX Calciun; Calcium-binding; SH3 domain.
SQ SEQUENCE 1097 AA; 120766 MW; 2573D7F8DCBACB CRC64;

Query Match 27.1%; Score 1702; DB 5; Length 1097;
Best Local Similarity 31.7%; Pred. No. 1.3e-77;
Matches 418; Conservative 203; Mismatches 329; Indels 368; Gaps 33;

11 SLIDVAIVYERAKDQGLSLKPTAGFTGQANPPFQSGLPQVLAQIMALDMMND 70
4 AVDAVAIVPRELTKYQEQFRALQPAQGFYGAQAQFLLQSLPPLIGQIMALDTSDD 63
71 GRMDVVSIAKMLKTLKLGQYQVLPSTLPVVKQGPVVAISSAPAGIGIASMP---LT 127
64 GKMNINSESIACKLNLKLRGMDVFKVLPSSL-----LSSL-----TGDDVSRMTRGSTS 113
128 AVAPVPMGSIPIVVGMSPLVSSVPAAVPPPLNNGAPVYIOLPFAHPAATWPKSSFSR 187
114 SLSPLD-----PLKGI-VPAVAVPVVAVPVA--VAATYISF-PGVSVSGPFPPTSN--- 162
138 SGFGSLQTLTKLQKQSPDVASAPPAEWAAPVPOSSSLKTRQLFNSHDKMTSGHLLTPQART 247
163 --PSPRHTSISERRAPSISSV---QGEWAVQAQKRYTQVFNANDRTSRGTLTSGQARG 217
248 ILMGSLPQAQALASIMNLSIDIDQGLKTAEEFILMHLIDVMSGOPLPVLPPEYIPPS 307
218 VLVQSKLPQVTLAQITWLTSDIDGRKLCNDEFILMLFCEKMAQSEKLPVTLPOEHWEPN 277
308 FRFVSGSGMSVSISSSVQDRLPEPPSSD-----EQQPEKLP--VFEDBK 353
278 LRKISRG---SVSGVSRSGSPASHASVSSQSGVGVADAPRTGLPQGTSEFEDXR 333
334 RENFSGVLEKRRQALIEQQRKEQRLAQLERAEQERKERERQEQAKRLBLEKOLE 413
334 KENYVGAQLDRRRKRKIMEDQQRKEERERKEREREADREKRALEAKRQEQELERLQ 393
414 KQRELERQREERREKRERERAKRELERQRLERERRRQRLNQRKKEQEGTVVLAR 473

DB 394 RQREIEMKEBQRRELEAEBAKRELEKQEQWEGARLAEMWAGREBERVLRKQKAH 453
QY 474 RKTLEFLEBALNDKQKOLEGLODIRLRLATQROEISTNKSRLRLAETHLQOQOES 533
DB 454 NTQNLVELSTLNKEIKELSORICPTRAGVNVKTVIDGMRTQRTSSEMSQLARKEQ 513
QY 534 QOMGRILPEK-----QILSDQLKVOQONSILRSDSLTLKRLALEKELELAROQLR 582
DB 514 NAKLQLTQERAKKEAKSKASGALGEMNAQEQ-----LNAFAHKOIINQIK 563
QY 583 EQLDEVERETRSLQOEIDVFNNOKELERE-----IHSKQOLQKRSLEARLKQEQERK 637
DB 564 DKVENISKEIESKEDINTNDVQMSLEKALSALITCEDLYKEYDVQRTSVLELKYNRK 623
QY 638 SLELEKQEDAKQRRVQERDKQWLEHVQOEOPRRKHEEDRLKREDSYKKEKEBAKP 697
DB 624 NETSVSAMDT-----GSSSNW-----EEGTITVDY-----AVASNDISALAAP 664
QY 698 EMQDKQRLFPHQEPAKLATQAPWSTTEKQPLTISAQESVYVYRALVPFESRSHDEI 757
DB 665 AVD-----LGSPAP-----EG-----FKYQAVIEFNARNAEII 693
QY 758 TIQPGDIWMVDESQGTGPGWLGELKGTGWFNPANYAEKIPENEVPTPAKPVTDLTSAFA 817
DB 694 TFFVGDIILVLEGNAPGWLAGEINGHTGWFPSYVEKLEGEV-----A 739
QY 818 PKALRETPAPLPVTSSEPTTPPNMADFSTWSSSNEKETNNMTMAAQLTYVPSA 877
DB 740 PVAAVE---APVDQAVTVADTVN-----DNINT-----SSIPAA 771
QY 878 GQLRQSAFPATATGSSPSPVLQGEKVEGLQAQALYPRAKKDNHLENKSDVITYLE 937
DB 772 S-----ADLTAG-----DV----- 781
QY 938 QQDMMWFGEVQGXGKWPFSKYVKLISGPVKSTSIDTPTSPASLKRVASPAKPAIPG 997
DB 782 ----- 781
QY 998 EEFIAMTYTESSEGGDLTFQGDVIVYTKDGMWTTGVDKSGVPSNRYRLKD----- 1052
DB 782 EYVIAAPYEAEBEGDLSFSAGEMWVYIKKEGEMWTTIGSRITGMPFSNRYQKADVGTAS 841
QY 1053 -----SESG----- 1057
DB 842 TAAAEVPSLDQETTLNGNAATYAPVAQEQVQIPVPQEPSPQPISSPGVGAEBAAHD 901
QY 1058 -----TAGKT-----GSLGKKPEIAQVYASYAATGPQOLT 1087
DB 902 LDTEVSGQNTQSKTQOSEPAESYSRPMSTRSSMTPGRAKRSLEIAQVIAPEATSTQOLS 961
QY 1088 LAPQQLIIRKKKNGPQWMEGELQARKKKRQIGMPANVYKLLSG--TSKTTPELAKTA 1145
DB 962 LTRQGLIMIRKKTGSGWMEGELQARKRRQIGMPANVYKLLSG--TSKTTPELAKTA 1021
QY 1146 VQPAVVC-QVIGMYVTQONDELAFSGQIINVNKEDPDWKGVEGQVGLPSPSNVY 1202
DB 1022 MTEQILDLKVALYLYKQONDELFSFDODIISVIGRDEPPEMRGELNGLSGLPSPSNVY 1079

RESULT 6
061618
ID 061618 PRELIMINARY; PRT; 1094 AA.
AC 061618;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Dynamn associated protein isoform DAP160-1.
GN DAP160 OR CGI099.
OS *Drosophila melanogaster* (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pezomyzeta;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; *Drosophilidae*; *Drosophila*.

NCBI_Taxid=7227;
 (1)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Head;
 RA Roos J., Kelly R.B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
 DR EMBL; AF053957; AAC39138.1; -
 DR HSSP; P29354; 1GR1.
 DR FlyBase; FBgn00233381; Dap160.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000261; EFS15_Homology.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00036; efhand; 2.
 DR Pfam; PF00018; SH3; 4.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 4.
 DR SMART; SM00054; EFh; 2.
 DR SMART; SM00027; EH; 2.
 DR SMART; SM00326; SH3; 4.
 DR PROSITE; PS00018; EF_HAND; 2.
 DR PROSITE; PS00031; EH; 2.
 DR PROSITE; PS00002; SH3; 4.
 KW SH3 domain.
 SQ SEQUENCE 1094 AA; 120495 MW; 141370EB1FB960F0 CRC64;

Query Match 27.1%; Score 1698.5; DB 5; Length 1094;
 Best Local Similarity 31.6%; Pred. No. 1.9e-77;
 Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;

11 SLDVAITVEERAKHDQFLSKPIAGFITGDQARNFPGSGLPQVLAQIWMALADNMND 70
 4 AYDAMAVTPREERAKYQEOFRALQPOAGFTGAQAKFFLQSQLPRILOQIWMALADTSD 63
 71 GEMDOVEFSIAMKLILKLTQGYQLPSTLPVWKQQVVAISSAPAFGIGIASMP---LT 127
 64 GKKNINEFSIACKLIMKLKGMVDVFKVLPSTL-----LSSL-----TGDPVSMTPGSGTS 113
 128 AVAPVPMGSIPIVWGSGPPLVSSVPPAAVPPPLANGAPVYQPLPAFPAHAPATPKSSFSR 187
 114 SLSPLD---PLKGI-VPAVAPVPPVAPVPA---VATVISP-FGVSVPFGPTPTPTSN--- 162
 188 SGPGLSNTKLQKQSFVDAASAPPAEMAVNPQSSRLKYROLFNSHDKTSGHLTGPQART 247
 163 --PSPHNTISERAPIESVN--QGMNAVQAQKKTQVFNANRRTSGHLSQANG 217
 248 IIMQSSLPQAQLASIWNLSDIDQDKLTAEFFILAMHLIDVAMSGQLPPLVPEEYIPPS 307
 218 VLVQSYLPQVTLAQIWTLSIDIDGRLNCDEFILAMFLCEKMAAGEKIPVTLPQEWVPPN 277
 308 FRRVSGSGMSVYSSSSVQRLPEPSSSD-----EQPEKLP--VTFEDK 353
 278 LKKIKSRPG---SVGVVSRPQSPASHASVSSQSGVGVADADTAGLPQGTSPEDK 333
 354 RENFERGVELEERKROALTEQOKKEOERLAOLERAOEKEREKROEAKROLELEKOLE 413
 334 KENYVQGAELDRKKIMEDQOKKEBEERKEEREADEKREKRLAEKQOELEKQOE 393
 414 KORELEKREERKEERKEEREAKELEBROLEWRNRROELNORNEOEGTVLTKAR 473
 394 ROREIEMERKEORKELEAKREKELERQOEWEQAIEMNAQKEREQEVVLKQXAH 453
 474 KRTLEFELALNDKQKQLESGKQDTCRLATROEIESTNKSRELIATITLQOOLQOS 533
 454 NTQNLVELSTLNEKIKELSORICDTAGVNTVTDGKRTORDTSSMSQKAIKIQO 513
 534 QOMAGRLIEPK-----QILSDQKVOYQONSILHRDLSLTLEKALEAKELANQQR 582
 514 NAKTLQLTORAKWEAKSKASGALGEMNAQOEQ-----LNAAPAHKQILINQK 563
 583 EQLDEVERETRSKQLEIDVFNQKLELR-----IHSKOOLOKORSLEAARLKQOEERK 637
 564 DKVENISKEISKEDINTNDVQMSLKAELSLALIKCEDLYKEVYVQVTSVLELKNK 623

638 SLELEKQEDARQVQERDKQWLEHVQOEQPRPKRPHBEDRLKREDSYKKEABERAKP 697
 624 NETSVSSAMDT-----GSSSAW-----BETGTTTDPY-----AASDISALAAAP 664
 698 EMQDKOSRLFHHPQEPKATLQAPWSTTEKGLPITISAOSVYVYRYALYPRESHDEI 757
 665 AYD-----LGGPAP-----EG-----FVYQAVYENNAABEI 693
 758 TIQPGDIYMWDESQIGEPGWLGGELKGTGMFPANYAEKIPENEPVTPAKPVTLTSDAP 817
 694 TEVPEDIIILVPLEQNAEPQWLAGIEINGHTGMFPESYVEKLEVEGV-----A 739
 818 PKLALRETPAPLPVYSSSEPTTPNNWADPSSSTWPESSSNEKEPTDWDMAQPSLTVP9A 877
 740 PVAAV-LEAVDAQVADT-----YNDNINT-----SSTIPA 768
 878 GQLRORSAFTPATATGSSPSPVLQGEKEVEGLQALYPMRAKDNHNFNKSVDITVLE 937
 769 S-----ADLTAAQ-----DY----- 778
 938 QODMMWFGEVQCGKMPKSYVKLISGVRKSTSIDTGPTEPSALKRVASPAKPAIPG 997
 779 ----- 778
 998 EEFIAMTYESESQGDLPQCGDVIATKKGDMWTGTGVGDKSGVFPSTNYVRKLD----- 1052
 779 EYVIAVPESEBBDLSFSAEWMVVIKKEGEMWTGTISRTGMFPSTNYVQKADVGTA 838
 1053 -----SEGSG----- 1057
 839 TAAAPVESLDQETTLNNGNAATAPVPAQEVQVPLVQERSBPISPGVGAEEAHED 898
 1058 -----TAGKT-----GSLGKKPEIAQVIAVTAATPEQTL 1087
 899 LDTEVSQINTQSKTQSSSEPAESYSRPMGRSTSMTPGMAKSEIAQVIAPEATSTQLS 958
 1088 LAPGQILIRKKNPGGMWEGELQANGKROIGMFPANTYKLSIPG--TSKITPTLEPKTA 1145
 959 LTRGQLIMIRKKTDSGMWEGELQANGRRRQIGWFPATYKVLQGGGRNSGRNTPVSGSRI 1018
 1146 VQPAVC-QVIGWYDVTQANDELAFPSKQIINVLNKEDPDMWKGESGOVGLFSPNYV 1202
 1019 MTEQLIDKVIATYKQANDDELSDKDDIISVLRDEPBMWKGELNLSGLFSPNYV 1076

RESULT 7
 Q81WH9 PRELIMINARY; PRT; 345 AA.
 AC Q81WH9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Similar to interectin 1 (SH3 domain protein) (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP TISSUE=uterus;
 RA Strusberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC039036; AAH39036.1; -
 FT NON TER 345
 SQ SEQUENCE 345 AA; 37077 MW; 967527FDE570DE85 CRC64;
 Query Match 26.9%; Score 1684; DB 4; Length 345;
 Best Local Similarity 95.3%; Pred. No. 2.5e-77;
 Matches 328; Conservative 6; Mismatches 10; Indels 0; Gaps 0;
 1 MAQFPTPGGSLDVVAITVEERAKHDQFLSKPIAGFITGDQARNFPGSGLPQVLAQ 60


```

Db      1 MAQFPTPGSGSLDVAITVEERAKHDQGFHSLKPISGFTIGDQANFFPGSLGPQVLAQ 60
Qy      61 IVALADNNNDGMDQVEFSIANKLTKLQGYQLPSTLPVWKQOPVAISSAPAFIGICI 120
Db      61 IVALADNNNDGMDQVEFSIANKLTKLQGYQLPSTLPVWKQOPVAISSAPAFIGICI 120
Qy      121 ASMPPLTAVAPVPMGSIPIVGVSPVSVPPAAPPVLANGAPVITQPLPAFAHAAWMP 180
Db      121 ASMPPLTAVAPVPMGSIPIVGVSPVSVPPAAPPVLANGAPVITQPLPAFAHAAWMP 180
Qy      181 KSSFSRSRSGPSQLTKLQKQSPDVASAPPAEAVPQSSRLKXROLFNHSHDKTMSGHL 240
Db      181 KSSFSRSRSGPSQLTKLQKQSPDVASAPPAEAVPQSSRLKXROLFNHSHDKTMSGHL 240
Qy      241 TGPQARTILMOSLSLQOQLASIMWISDIDQDQKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Db      241 TGPQARTILMOSLSLQOQLASIMWISDIDQDQKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Qy      301 PEYIPFRRVRVSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344
Db      301 PEYIPFRRVRVSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344

```

RESULT 8

```

ID      081W15      PRELIMINARY;      PRT;      611 AA.
AC      081W15;
DT      01-MAR-2003 (Tremblrel. 23, Created)
DT      01-MAR-2003 (Tremblrel. 23, Last sequence update)
DE      01-MAR-2003 (Tremblrel. 23, Last annotation update)
OS      Similar to intersecin 2 (Fragment).
OC      Homo sapiens (Human)
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      Tissue=Skin;
RA      Strubberg R.;
RL      Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.
DR      EMBL; BC038963; AAH38963.1; -.
FT      NON TER      611
SQ      SEQUENCE      611 AA; 69236 MW; 85F6EA0C664A881A CRC64;

```

Query Match 24.4%; Score 1330; DB 4; Length 611;
 Best Local Similarity 54.0%; Pred. No. 2,8e-69;
 Matches 331; Conservative 86; Mismatches 118; Indels 78; Gaps 14;

```

Qy      1 MAQFPTPGSGSLDVAITVEERAKHDQGFHSLKPIAGFTIGDQANFFPGSLGPQVLAQ 60
Db      1 MAQFPTPGSGSLDVAITVEERAKHDQGFHSLKPIAGFTIGDQANFFPGSLGPQVLAQ 60
Qy      61 IVALADNNNDGMDQVEFSIANKLTKLQGYQLPSTLPVWKQOPV-AISSAPAFIGICI 118
Db      61 IVALADNNNDGMDQVEFSIANKLTKLQGYQLPSTLPVWKQOPV-AISSAPAFIGICI 118
Qy      119 GIASMPPLTAVAPVPMGSIPIVGVSPVSVPPAAPPVLANGAPV 165
Db      119 GIASMPPLTAVAPVPMGSIPIVGVSPVSVPPAAPPVLANGAPV 165
Qy      166 IOPPLAFAPATWPKSSSFS--RSQPSQLTKLQKQSPDVASAPPAEAVPQSSRLKXROLFNHSHDKTMSGHL 242
Db      166 IOPPLAFAPATWPKSSSFS--RSQPSQLTKLQKQSPDVASAPPAEAVPQSSRLKXROLFNHSHDKTMSGHL 242
Qy      243 POARTILMOSLSLQOQLASIMWISDIDQDQKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Db      243 POARTILMOSLSLQOQLASIMWISDIDQDQKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Qy      301 PEYIPFRRVRVSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344
Db      301 PEYIPFRRVRVSGSGSVISSVSDQRLPEEPPSEDEQOPEKK 344

```

```

Db      350 LVPSFR-----GKQI---DSINGTLPSPYQKQ-EEEPQKPLPTVEDKRYANERGM 400
Qy      363 ELEKRRQALLBQKRECEQERLQLERAQOEKREKEROEAKQOLEKOLEKORELOR 422
Db      401 ELEKRRQALLBQKRECEQERLQLERAQOEKREKEROEAKQOLEKOLEKORELOR 460
Qy      423 EEEERKTEEREAQRLEEROROLEMERNRROELLNORKEOEGTVLKARKTLEFELE 482
Db      461 EEEERKTEEREAQRLEEROROLEMERNRROELLNORKEOEGTVLKARKTLEFELE 520
Qy      483 ALNDKQLBEGKLDIDICRLATQROEISTNKSRELIAETIHLQOQLOESQOMGLIIP 542
Db      521 ALNDKQLBEGKLDIDICRLATQROEISTNKSRELIAETIHLQOQLOESQOMGLIIP 580
Qy      543 EKQILSDQKQV 555
Db      581 EKQILSDQKQV 593

```

RESULT 9

```

ID      09U2T9      PRELIMINARY;      PRT;      1097 AA.
AC      09U2T9;
DT      01-MAY-2000 (Tremblrel. 13, Created)
DT      01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE      01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE      Y116A8C.36 protein.
GN      Y116A8C.36.
OS      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Peloderiinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      McMurtry A.A.;
RL      Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99069613; PubMed=9851916;
RA      none;
RT      "Genome sequence of the nematode C. elegans: A platform for
RT      investigating biology";
RT      Science 282:2012-2018 (1998).
CC      -i- SIMILARITY: CONTRAINS 5 SH3 DOMAINS.
DR      EMBL; AL117204; CAB5138.1; -.
DR      HSSP; P29355; ISEM.
DR      WormPep; Y116A8C.36; CE23342.
DR      InterPro; IPR000194; ATPase a/bcentre.
DR      InterPro; IPR002048; EF-hand.
DR      InterPro; IPR000261; ERS15_homology.
DR      InterPro; IPR001452; SH3.
DR      Pfam; PF00036; ethand; 2.
DR      Pfam; PF00018; SH3; 5.
DR      PRINTS; PR00452; SH3DOMAIN.
DR      ProDom; PD000666; SH3; 5.
DR      SMART; SM00027; EH; 2.
DR      SMART; SM00326; SH3; 5.
DR      PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR      PROSITE; PS00018; EF_HAND; 2.
DR      PROSITE; PS50031; EH; 2.
DR      PROSITE; PS50002; SH3; 5.
KM      SH3 domain.
SQ      SEQUENCE      1097 AA; 122072 MW; 1C2BA5F103968372 CRC64;

```

Query Match 21.3%; Score 1333; DB 5; Length 1097;
 Best Local Similarity 29.4%; Pred. No. 4,8e-59;
 Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;

```

Qy      15 WAITVEERAKHDQGFHSLKPIAGFTIGDQ-----ANFFPGSLGPQVLAQIIVALADM 67
Db      5 WEVSDAEYQKFAF-----GQLTGQGPMDAVTARNALMRNLPTQVLSQIIVALSDL 57

```

QY	68	NNDDGMDQVPSIAKKLLKTLQOGQLTSTLPVWKQOPVAISSAPAGIGIASMPLT	127
Db	58	DDGGLDIREISIAIRLLNLCTAGIPITPOJPSLLVVA--RNAPTWPS-----	107
QY	128	AVAPPMGSIPIVWGMSPLV-----SSVPPAAVPLNCGAPVITQIPAPAHPATWPK	181
Db	108	-----RHSGSVSYQLPPAIDRMSQSIIPSA--PVSIAGTP-----	142
QY	182	SSSFERSGGSQLNTKLOKASFDVASAPPAEWAVPOSSRLKTRQLFNSHDKTWSGHLT	241
Db	143	SSRHNSISAGSELN-----NDRNVEGRQOLEMNAI PHHNKLYKSQLFNALDKERLGSLT	196
QY	242	GPQATILMOSSLPOAQLASTIMNLSIDODGLTLEEFITLMLHLIDVMSQPLPVLP	301
Db	197	SOVGSALGSLPLFNVLAHITWFSIDVKKDGLSDEVAISQYMIEMKSGFAPAKTIP	256
QY	302	EYIPSPFRVRSQMSVYSSSSVQORLPPEPSSEDEOPEKULPVTFEDKKRENFERS	361
Db	257	ELV-----RMGSISSRSANNTLELPGAPPO--KSPAKTIEDKQDNLSKQ	303
QY	362	VELEKROALLEQOREQERLAOLERAEOERKEREROEAKROLELEKOLEKOLERO	421
Db	304	AELERRRHLVEEEOGRRAVEKKEKEBEAKNRRQEKQEAVERQAELEBRQIILEAQ	363
QY	422	REEE---RKLEER-----EAKRELE---ROQLEWERNRROELLNQRNKEOEGTV	469
Db	364	REEEKKRLEMERREDEDEKKRKQOMEKAKYQOMVPEQKKNFYNOQOMENERLAQ	423
QY	470	LKARKTLEFELALNDKKHOLEGLDORCLATQROEITSNKSRBLTIAETIHQQQ	529
Db	424	RQORRTIQLQOLADKXYIDVEVDIGAKAEVAEVTGFIEMSTRDEKARI---KE	479
QY	530	LOESQOMLGRLLPEKQILSDLOKYOONSLRHSDLSLLTKRALAEKELARQOLRELD--	586
Db	480	LOETNQ---KTAISEQELGHOL--LOQSAHMET--TQKSELEALRRKDAIRALIEDPA	533
QY	587	-EVERETSKLOEIDVF--NOLUKELREHKSQOLOKRSLEAPRLKQKEGRKSLELEKQ	644
Db	534	LELSTEKESYNQTEILTKTNEKYTDDYSK-----LVAKREERYNSFEL---	578
QY	645	KEDAGRVOERDKQWLEHVOOEORPRKP-----HEEDLKREDSVRK	688
Db	579	-----LVHAQTHARSKTIGFEAKSAPASAAPAPAPAPATTNGPPANNDPGEFD--K	631
QY	689	KEAERAPREMOKOSRLFHPHOEPAKLATQAPWSTTEKGLTISAGESVAVVYRALYP	748
Db	632	TDASGRPADGATST-----ADPFAQIQAAP--AHSKGAVDQSAFNIHTYKCRALFA	683
QY	749	FESRSHDEITTOPGDIVWDSQTCGPEWLGELKGTGMFPYANAEKIPENEVPTPAKP	808
Db	684	FEARSEDELSFEPGDVIVIFOSHAAPEPMRAGOLEKYGWPEAPFEVAIA--AVTPG--	739
QY	809	VTDLSAPAKLALRETPAPLPVYSSESTTPNNMADSSITWPPSSSNKKEPTDMMWTMA	868
Db	740	-----GDPIQOMPNNMTPTSSVDDIGV--KAARKAI-----AA	772
QY	869	QSLTVPAGOLRQSAFTPATATGSSSPVLQGEKEVGOAOLYPMRAKKONHNLNPN	928
Db	773	AMGLTEGGA-----PRASSAPAAVAIS-----QCLAQPMKRAAREDELSFA	814
QY	929	KSDVTVLEQODMMWF--EVOGQGMKFPKSYVKLISGFVKRSTSIDGPTESPASLKVA	987
Db	815	KGDTLEVELEKEMKMKGNRPAGEITGMFPKSYKVEGATTTSTTTP--VSPKASAGAPOAA	873
QY	988	SPAAPKALPGEF-----FIAMTYESEGGDLFFQOGDVIWTKKQGDWWTG	1034
Db	874	AGAQYDVPSDVTTLQASSTAPQOOLYTVIYDEAVETDILALHVGDTLLVLEKDEWKG	933
QY	1035	TVGDSGVFPSPNVYALKDSGSGTAGTK--GSLGKPEI---AGVISAATGCEOLITLA	1089
Db	934	RONGREGIFPANYVEISVQAGDPIRPPQATPPAPAPVVLCEAKYVUDVFVASANQLOJIK	993
QY	1090	PGQLILKKKPGCWMBEGELQARGKKROIGMFPANYVLLSPGTSKITPTLEPKTAVQPA	1149

DB	Query Match	Best Local Similarity	Matches	21.0%; Score 1314; DB 11; Length 755; Pred. No. 2.7e-58; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB	994 VGETVTKIREKSAAQMWGEGLIRNGKP-IAQMPFGEYVKVILEAASPAT-----	1040		
QY	1150 VCOYIGMWADYIAQNDDELAFSKQGIINVLNKEDPDMWKG---EVSQGVGLFPSNYVK	1203		
DB	1041 --RATAYVDYFASQPDDELGFRTGQVITVTBKSLAEFMWSGIREQDPSKGLFPSNYVQ	1095		
RESULT 10				
ID	08C4B5	PRELIMINARY;	PRT;	755 AA.
AC	08C4B5			
DT	01-MAR-2003 (Tremblrel. 23, Created)			
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	Intersectin (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RL	60,770 full-length cDNAs";			
RL	Nature 420:563-573(2002)";			
DR	EMBL, AK082606; BAC38546.1; -.			
FT	NON TER			
SO	SEQUENCE	755 AA; 85107 MW; 8FD0026865297597 CRC64;		
QY	964 GPRVKSSTIDGTPESPASLKRVASPAKCAIPEEFIAMYTYESSHOGDLTFQOGDVIY	1023		
DB	5 GPRVKSSTIDGTPESPASLKRVASPAKCAIPEEFIAMYTYESSHOGDLTFQOGDVIY	64		
QY	1024 VTKKDGWMVTGTVDKSGVFPSPNYVRLKDSESGGTAGKTSLGKKPEIAVIASVYATGP	1083		
DB	65 VTKKDGWMVTGTVDKSGVFPSPNYVRLKDSESGGTAGKTSLGKKPEIAVIASVYATGP	124		
QY	1084 EQLTLAAGQILIRKKNPGMWGEGLQARKKRQIGFPPANNYKLSPGTSKTTPELPK	1143		
DB	125 EQLTLAAGQILIRKKNPGMWGEGLQARKKRQIGFPPANNYKLSPGTSKTTPELPK	184		
QY	1144 TAVOPAVCOYIGMWADYIAQNDDELAFSKQGIINVLNKEDPDMWKGEVSGGVGLFPSNYVK	1203		
DB	185 TAVOPAVCOYIGMWADYIAQNDDELAFSKQGIINVLNKEDPDMWKGEVSGGVGLFPSNYVK	244		
QY	1204 LTTDMDPSSQ 1213			
DB	245 LTTDMDPSSQ 254			
RESULT 11				
ID	08C9C3	PRELIMINARY;	PRT;	440 AA.
AC	08C9C3			
DT	01-MAR-2003 (Tremblrel. 23, Created)			
DT	01-MAR-2003 (Tremblrel. 23, Last sequence update)			
DT	01-MAR-2003 (Tremblrel. 23, Last annotation update)			
DE	SH3 domain protein 1B (Fragment).			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Thymus;			

RX MEDLINE=22354683; PubMed=12466851;
 RA The RANTOM Consortium;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK044449; BAC31264.1; -
 FT NON TER 440
 SQ SEQUENCE 440 AA; 48760 MW; 1C94A5E62B71EEF2 CRC64;

Query Match 18.7%; Score 1172; DB 11; Length 440;
 Best Local Similarity 56.7%; Pred. No. 2e-51;
 Matches 261; Conservative 58; Mismatches 87; Indels 54; Gaps 15;

QY 1 MAQFTPPGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 DB 1 MAQFTPPGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 QY 61 IVALADNNDRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--AISAPAFGIG 118
 DB 61 IVALADNNDRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--AISAPAFGIG 118
 QY 119 GIASNP-----PLTAVAPV--PMG-----SIPVAGSPPLVSSVPPAAPPPLANGAP 164
 DB 120 ---SNPNLSIHQPLPVPAPVPIATPLSSAGTSTPLPMAPPLVPSVTSLSLP---NGTAS 173
 QY 165 VIQPLPAHPAATPKSSFS--RSGFGSLNTKLOKQKQ--FDVASAP----- 211
 DB 174 LIQPL-SIPYSSSTLPHASYSYSLMNGFG--GASIQKASLIDIDSSSTSTASLSGN 229
 QY 212 -----AAEVAPOSRLKRYROLFNHSDTKMSGLTGPQARTIMOSLPQOLASIMNLS 266
 DB 220 SPKGTSTSEKAVDPQRLKRYROLFNHSDTKMSGLTGPQARTIMOSLPQOLASIMNLS 289
 QY 267 DIDQDKLTAEFFILAMHLIDVAMSGQLPVLPEPVIIPSEFRVRSAGMSVSSSSVD 326
 DB 290 DIDGQGLKAEFFILAMHLIDVAMSGQLPVLPEPVIIPSEFRVRSAGMSVSSSSVD 341
 QY 327 QRLPEPSSSEDEQOEKLPVTFEDKRENFGRSVLEKROALLBOORKEQLAQL 386
 DB 342 GTLPYQKIQ--BEEPOKLPVTFEDKRENFGRSVLEKROALLBOORKEQLAQL 400
 QY 367 RAEQKERERQOEAKROLLEKOLEKORELEROREER 426
 DB 401 KEWERKQRELOEQEKOLEKOLEKORELEROREER 440

RESULT 12

O8CJ43 PRELIMINARY; PRT; 164 AA.

AC O8CJ43; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Intersectin isoform 4 (Fragment).
 GN ITSN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NX NCBI_TaxId=10090;
 RX SEQUENCE FROM N.A.
 RC STRAIN=BAIB/cx57BLP1; TISSUE=lung;
 RA Skripkina I.Ya., Teyba L.A., Slavov D., Gardiner K., Rynditch A.V.;
 RT "Novel splicing forms of human and mouse Intersectin gene
 transcripts."
 RL Ukr. Biokhim. Zh. 74:33-43(2002).
 DR EMBL; AF525079; AAAT5695.1; -
 FT NON TER 1
 SQ SEQUENCE 164 AA; 17677 MW; 5E5868CA919DE969 CRC64;

Query Match 13.1%; Score 822.5; DB 11; Length 164;

Best Local Similarity 81.6%; Pred. No. 2.3e-34;
 Matches 164; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 2 AQPFTPPGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 61
 DB 1 AQPFTPPGSLDVAITVEERAKDQFLSKPIAGITGDQANFFFGSLPPVLAQ 60
 QY 62 WALADNNDRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--AISAPAFGIG 121
 DB 61 WALADNNDRMDQVERSIAMKLIKLOGLPSTLPVWKQOPV--AISAPAFGIG 114
 QY 122 SMPPLTAVAPVPMGISIPVAGSPPLVSSVPPAAPPPLANGAPPVIOPLPAFAPATWPK 181
 DB 115 -----AAVPPLANGAPPVIOPLPAFAPATWPK 143
 QY 182 SSSFSRSGPSQLNTKLOKQKQ 202
 DB 144 SSSFSRSGPSQLNTKLOKQKQ 164

RESULT 13

O8CGU5 PRELIMINARY; PRT; 158 AA.

AC O8CGU5; 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Intersectin isoform 5 (Fragment).
 GN ITSN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NX NCBI_TaxId=10090;
 RX SEQUENCE FROM N.A.
 RC STRAIN=BAIB/cx57BLP1; TISSUE=Brain;
 RA Skripkina I.Ya., Teyba L.A., Slavov D., Gardiner K., Rynditch A.V.;
 RT "Novel splicing forms of human and mouse Intersectin gene
 transcripts."
 RL Ukr. Biokhim. Zh. 74:33-43(2002).
 DR EMBL; AY127576; AA02285.1; -
 FT NON TER 1
 SQ SEQUENCE 158 AA; 17252 MW; 4DD3D8D070CC4E5 CRC64;

Query Match 13.0%; Score 812.5; DB 11; Length 158;
 Best Local Similarity 96.8%; Pred. No. 7.1e-34;
 Matches 153; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 708 HHQEPKLTQAQWSTTEKGPLTISAQESKVVYRYALVPFESRSHDEITIOGDIVM- 766
 DB 1 HHQEPKLTQAQWSTTEKGPLTISAQESKVVYRYALVPFESRSHDEITIOGDIVM 60
 QY 767 -----VDESQGEPMGLGELKGTGWFPANYAEKIPENVEVTPAKPYTDLTSAPAKLAL 822
 DB 61 KGEWVDESQGEPMGLGELKGTGWFPANYAEKIPENVEVTPAKPYTDLTSAPAKLAL 120
 QY 823 RETPAPLPVTISSEPTTPNNWADPSSSTWPSSSNEKPEP 860
 DB 121 RETPAPLPVTISSEPTTPNNWADPSSSTWPSSSNEKPEP 158

RESULT 14

O8T068 PRELIMINARY; PRT; 662 AA.

AC O8T068; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE Intersectin isoform 23, Last annotation update)
 GN DAPI60 OR CG1099.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxId=7227;
 RN [1]
 RC STRAIN=Beerley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Chapeau M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Gelink S.,
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
 DR EMBL: AY069517; AL39662.1; -
 DR Flybase: FBgn0023388; Dap160.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 4.
 DR ProDom: PD000066; SH3; 4.
 DR SMART: SM00326; SH3; 4.
 DR PROSITE: PSS0002; SH3; 4.
 KW SH3 domain.
 SQ SEQUENCE 662 AA; 72553 MW; 3532CB7034B7D7AC CRC64;
 Query Match 12.7%; Score 794.5; DB 5; Length 662;
 Best Local Similarity 25.9%; Pred. No. 3.3e-32;
 Matches 221; Conservative 108; Mismatches 203; Indels 321; Gaps 21;
 QY 459 QNNKQEGIVLAKARKITLFELEALNDKQLEGLODIRLATORGEISTNKSREL 518
 DB 4 QKEROERLAKQKHNTOINLVELSTINEKELSORICTRAGVTVNKTVIDGMRTQRTD 63
 QY 519 RLAETITLQOOLQOESQWGLRIPEK-----QLTSQLKQOVQNSLHRSLLTL 567
 DB 64 SMSSESQLKRIKEQNAKLQLOERAKKAKSKSGAALGSGNAQOEO-----L 113
 QY 568 KALEAKELAQOLREQLDEVERETSKLQEIIVFNNOULKELE-----IHSKOLOKOR 622
 DB 114 NAAFAHKQIINQIKOVENISKEISKEDINTNDVQSELKAEALITCEDELYKEY 173
 QY 623 SLEAARLKQEKQERKSLLEKQEDAKQKQVQERDKOMLEHVOOEPKRRKHEEDRLK 682
 DB 174 DVQRTSVLELKYRNKETSVSASAMD-----GSSSNM-----ERTTIVTDY----- 216
 QY 663 EDVSATKEAEERAKEMODKQSLRPHQEPAKLATQAPWSTTEKGLTISAQESKVVY 742
 DB 217 --AVASNDLSALAAPAVD-----LGGPAP-----EG-----FKK 243
 QY 743 YRALYPFESRSHDEITTOPGDIWMVDESQTEGPGWLGELKGTGWFPAVYAEKIPENEV 802
 DB 244 YQAYVEFNARNAEITTFVPGDIILVPLBQNAEPGMLAGEINHTGWFPSYVEKLEVGCV 303
 QY 803 PTPAPKVTDLTSPAPKALALRETPAPLPVTSSEPTTPNNMADFSTWSSNKEPETDN 862
 DB 304 -----LPVAVAVE-----APVDAQVATVADTVN-----DN 327
 QY 863 WDTPAAQPSLTVPASGLQKORSAFTPATGTSSPSVVLGGKREGLQALYPRAKKD 922
 DB 328 INT-----SSLPAS-----ADLTAG----- 344
 QY 923 NHIATFKSDVITVLEQDMMWFGEVQKQWFPKSYVKLISGPVKSIDSIDGTSPAS 982
 DB 345 -----DV----- 346
 QY 983 LKRVASPAKPAIPGEFTAMTYTSSSEGDLTFQOGDVIYVTKDGMWMTGTVDKSGV 1042
 DB 347 -----EYVIAAYPYBSABEGDLSFSAGEMWVYIKKEGEMWTGTISRTGM 391
 QY 1043 PPSNVYRLKD----- 1052
 DB 392 PPSNVYQKADVGTAATAAEVPESLDQETLLNGNAAYTAAPVBAEQVYQPLPVQBPSEQ 451
 QY 1053 ---SESGG-----TAGKT-----GSLGKKPEIA 1072

DB 452 PISSPGVGAEEAHEDLDTEVSQINTQSTQSSSEPAIYSRPMSTRSMTPGKARKESEA 511
 QY 1073 OVIASTATGPEQLTLARQOLILTRKNPGCWMEBELQARKKROIGFPANVYVLLSPG 1132
 DB 512 OVIAPEATSTEQSLTRGQOLIMIRKKTDSGWMEBELQAKGRRRQIGFPAVYVVLQGG 571
 QY 1133 --TSKITPELEPKTVQPAVC-QVIGMYDYTAQNDDELAFSGQIINVLNKEDPPMWKGE 1189
 DB 572 RNSGNNTVSGSRIMTQIIDLKIALYPRARQNDDELSPDKDILISVGRDEPEWMHGE 631
 QY 1190 VSGQVGLFPPSNV 1202
 DB 632 INGLSGLFPPSNV 644
 RESULT 15
 ID Q8CD59 PRELIMINARY; PRT; 276 AA.
 AC Q8CD59;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE SH3 domain protein 1B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573 (2002).
 DR EMBL: AK031390; BAC27381.1; -
 SQ SEQUENCE 276 AA; 33547 MW; BBEC93085EDCD27D CRC64;
 Query Match 11.2%; Score 704; DB 11; Length 276;
 Best Local Similarity 58.9%; Pred. No. 4e-28;
 Matches 145; Conservative 43; Mismatches 58; Indels 0; Gaps 0;
 QY 362 VLEKRRQALLQQRKEGERLQLERAQEKERERQEQEAKQOLEKQELERQ 421
 DB 1 MELEKRRQVLMQQRERAKRAQEKERERQERLOEQEWKQOLEKREKQELERQ 60
 QY 422 REEERKKEITEREAKRLEEROROLEMERNRROELLNORKEOEGTVLAKARKTLEPEL 481
 DB 61 REEERKKEITEREAKQOLEERORLEMERLROELLSQKREQEDIVALSRSKSLHLEL 120
 QY 482 EALNDKQKQLEGKLODIRCRLATORGEISTNKSRELRIAEITHLQOOLQOESQWGLRI 541
 DB 121 EAVNGHQQISGRLODVQIRKQOTKTELEVLDKQODLEIMEIKQOQGLKEVQNNLIYLV 180
 QY 542 PEKQLSQOLQOVQNSLHRSLLTLKALAEKELARQOLREQLDEVERETRSKIQEIDV 601
 DB 181 PEKQILNRIKQMSLNTPDGSIHLKKSSEKELCQRLKEQDALKETASKLSEWDS 240
 QY 602 FNNQLK 607
 DB 241 FNNQLK 246
 Search completed: December 4, 2003, 15:20:09
 Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 21:34:26 ; Search time 770.054 Seconds
(without alignments)
15719.093 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642
Sequence: 1 atggctcagttccaccaccc.....tgaccaccagccagcaatga 3642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first: 45 summaries

Database: Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	477.4	13.1	2017	10	US-09-884-441-72	Sequence 72, App
2	477.4	13.1	2017	11	US-09-907-969-72	Sequence 72, App
3	477.4	13.1	2017	12	US-09-827-271-72	Sequence 72, App
4	477.4	13.1	2017	14	US-10-198-053-72	Sequence 72, App
5	356	9.8	568	11	US-09-764-881-55	Sequence 55, App
6	332.4	9.1	2873	9	US-09-879-957-193	Sequence 193, App
7	331.4	9.1	4210	10	US-09-764-868-125	Sequence 125, App
8	322.6	8.9	503	11	US-09-918-995-31258	Sequence 31258, App
9	258	7.1	747	9	US-09-879-957-39	Sequence 39, App1
10	198.8	5.5	270	9	US-09-864-761-17127	Sequence 17127, A
11	198.8	5.5	286	9	US-09-864-761-26948	Sequence 26948, A
12	198.8	5.5	297	9	US-09-864-761-30453	Sequence 30453, A
13	198.6	5.5	301	9	US-09-864-761-17146	Sequence 17146, A
14	197.4	5.4	263	9	US-09-864-761-17644	Sequence 17644, A
15	192	5.3	480	9	US-09-864-761-10314	Sequence 10314, A
16	180.8	5.0	486	9	US-09-864-761-333	Sequence 333, App

17	180.8	5.0	487	9	US-09-864-761-864	Sequence 864, App
18	163	4.5	475	9	US-09-864-761-311	Sequence 311, App
19	163	4.5	475	9	US-09-864-761-13884	Sequence 13884, A
20	154.6	4.2	967	11	US-09-764-881-50	Sequence 50, Appl
21	146.4	4.0	180	9	US-09-864-761-17125	Sequence 17125, A
22	144.2	4.0	292	9	US-09-864-761-20261	Sequence 20261, A
23	144.2	4.0	304	9	US-09-864-761-21373	Sequence 21373, A
24	144.2	4.0	310	9	US-09-864-761-19751	Sequence 19751, A
25	144.2	4.0	310	9	US-09-864-761-19759	Sequence 19759, A
26	144	4.0	480	10	US-09-884-441-60	Sequence 60, Appl
27	144	4.0	480	11	US-09-907-969-60	Sequence 60, Appl
28	144	4.0	480	12	US-09-827-271-60	Sequence 60, Appl
29	144	4.0	480	14	US-10-198-053-60	Sequence 60, Appl
30	144	4.0	531	10	US-09-884-441-5	Sequence 5, Appl
31	144	4.0	531	11	US-09-907-969-5	Sequence 5, Appl
32	144	4.0	531	12	US-09-827-271-5	Sequence 5, Appl
33	144	4.0	531	14	US-10-198-053-5	Sequence 5, Appl
34	144	4.0	3489	12	US-10-294-804-1	Sequence 1, Appl
35	142.4	3.9	955	11	US-09-764-881-54	Sequence 54, Appl
36	140.4	3.9	230	9	US-09-864-761-17643	Sequence 17643, A
37	140.4	3.9	247	9	US-09-864-761-30501	Sequence 30501, A
38	132.2	3.6	462	11	US-09-918-995-21728	Sequence 21728, A
39	127	3.5	419	9	US-09-864-761-309	Sequence 309, App
40	125.2	3.4	484	9	US-09-864-761-863	Sequence 863, App
41	121.6	3.3	531	10	US-09-884-441-4	Sequence 4, Appl
42	121.6	3.3	531	11	US-09-907-969-4	Sequence 4, Appl
43	121.6	3.3	531	12	US-09-827-271-4	Sequence 4, Appl
44	121.6	3.3	531	14	US-10-198-053-4	Sequence 4, Appl
45	119.2	3.3	1926	12	US-10-294-804-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1	US-09-884-441-72
1	Sequence 72, Application US/09884441
2	Patent No. US20020119158A1
3	GENERAL INFORMATION:
4	APPLICANT: Algate, Paul A.
5	TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
6	FILE REFERENCE: 210121.462C7
7	CURRENT APPLICATION NUMBER: US/09/884,441
8	NUMBER OF SEQ ID NOS: 489
9	SOFTWARE: FastSeq for Windows Version 3.0
10	SEQ ID NO 72
11	LENGTH: 2017
12	TYPE: DNA
13	ORGANISM: Homo sapien
14	US-09-884-441-72
15	Query Match
16	Best Local Similarity 13.1%; Score 477.4; DB 10; Length 2017;
17	Matches 111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;
18	1 ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG 60
19	42 ATGGCTCAGTTTCCACACCTTTCGTTGCTGATGCTGGGCCATCTGTGAG 101
20	61 GAAAGGGCCAGATGACGAGATGCTTACCTGAGCCGATGAGGGATTTATTCT 120
21	102 GAACGATCTACACATGATTAACAGTTTGTATTAACCTTCAAGAGGTTTACATTA 161
22	121 GGTGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
23	162 GGTGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 221
24	181 ATATGGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
25	222 ATATGGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 281

```

OY 241 GCCATGAGCTTATGAACTGAGCTACAGATATGAGCTCCCTCCACTTCCCT 300
DB 282 GCTATGAACCTATCAATTAAGTTGAGGCGCAAGCTGCTGATGCTCTCTCT 341
OY 301 GTCATGAACAGCAACCACTG---GCTATTTCCAGTGCACCACTTTGTATAGAGG 357
DB 342 ATCATGAACCAACCCCTATGCTCTCTCACTAATCTCTGCTCGTTTGGATGGAAAG 401
OY 358 ATTGCTAGCATG-----CCACACTCAACAGCTGTTGCTCTGTCGCA----- 400
DB 402 ATGCCCATCTGTCATTCATCAGCACTTGCCTCAGTTGCACTATAGCAACCCCTTG 461
OY 401 -----TGGCTCCATTGCAAGTTGTTGGAATGTCTCCACCCCTTGTATCT 444
DB 462 TCTTCTGCTACTTCAGGACCAAGTATCTCTCCCTAATGATGCTGCTCTGCTGCT 521
OY 445 TCTGTCCTCCAGCAGAGTGCCTCCCTGCTAAGGGGCTCTCCCTGATACAGCT 504
DB 522 TCTGTTAGTACATCTCTATTAACCAATGGAATGCGCATCTCATTCAGCCTTATCCAT 581
OY 505 CTGCTGCGTTTGGCATCTGCGACCAATGCGCAAAAGTTCTTCTTCAAGCATCT 564
DB 582 CTTATTTCTTCTTCAACATGCTCATCATCTTACAGCTGATGATGAGGAGATT 641
OY 565 GGTCCAGGCTCACTTAACCTAAGTTACAGAGCAACATCTTGTGATGCTGCCAGC 624
DB 642 GGTGATGCTAGTATCAGAAAGGCCAGTCTGATGATGATTAAGATCTAGTACTCACT 701
OY 625 GCGCTCCAG-----CAGCAAAATGGGCTGTG 651
DB 702 TCTCTAATCTGCTCCCTCTCAGAGGAATCACCCTAAGACAGGACCTCAGAGTGGGCA 761
OY 652 CTTGATCATCAGGCTGAATATACAGGAGTATTAACAAGCAGACAAACATATGAGT 711
DB 722 CTTGAGCTTCAAGATTAAGTACGCAAAATTAATAGCTAGACAAAGGCTAGAGC 821
OY 712 GGAACCTTAACAGGCTCCAGGCAAGATATCTCAAGCAATCAAGTTTACCCAGCT 771
DB 822 GATACCTCTCAGGTTTCAAGCTAAGAAATGCTCTTCACTCAAACTCTCTCAAACT 881
OY 772 CAGTGGCTTCAATATGAACTTTCTGACATTTGATCAAGATGAAATCACTGTCAGAA 831
DB 882 CAGCTAGCTACTATTTGAGCTCTGCTGATCATGATGATGAGAGCACTGTAAGCTGA 941
OY 832 GAATTTATCTAGCTATGACCTAATGATGTTGCCATGCTGCTGACGCACTGCCGCC 891
DB 942 GAATTTATCTGCGATGACCTCACTGACATGCGCAAGCTGGAACGCACTACCATG 1001
OY 892 GTCCGCTCCAGAAATACATCCCTCTCTTCAAGAAAGTCCGCTCCGCGAGTGGAGT 951
DB 1002 ACGTGGCTCCGAGCTGTCTCTCATCTTCAAGAGGGGAAACAG----- 1050
OY 952 TCCGTCATAGCTCTTCTTCTGATGATCAGAGCTGCTGAGAGCCGCTGTCAGAGAT 1011
DB 1051 -----TTGATTTCTTTAATGAACTCTGCTTCAATACAGAAAAACAGAA 1097
OY 1012 GAGCAGCAGCAGAGAAAGAACTGCTGATTTGAATTTGAATTAAGAGCGGAGATTC 1071
DB 1098 GA---AGAGCTCAGAAAGAACTGCCAGTTACTTTTGAAGACAAACGGAAGCCATAT 1154
OY 1072 GAGCGAGAGAGTGTGAGAGTGAAGAGCGCGCAAGCGCTTTGAGACAGAGAGGCAAA 1131
DB 1155 GAACGAGAAACATGAGAGTGAAGAGCGACGCAAGTGTGATGATGAGAGAGAGAGG 1214
OY 1132 GAGCAGAGAGGTTGCTCAGCTGAGCGCGCAGCAGAGAGAGAGAGAGCGGAGGCG 1191
DB 1215 GAGGCTGAACGCAAAAGCCAGAAAGAGAGAGAGAGTGGAGCGGAAACAGAGAACTG 1274
OY 1192 CAGGAGCAGAGAGCGCAAGCGGAGCTGAGCTGAGAGAGCACTGAGAGAGCGGAGG 1251
DB 1275 CAGAGCAAGATGAGAGAGAGAGCTGAGGTTGAGAGAAACGCTTGAAGAAACAGAGAG 1334

```

```

OY 1252 CTGAGCGCGAGCGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAA 1311
DB 1335 CTGAGAGACAGCGCGAGAGAGAGAGAGAGATTAAGAAAGCAGAGAGCAAAA 1394
OY 1312 CGGAACTGAAAGGCGAGCAACATTTGAATGGAAACGAAACCGAGACAGAACTCTG 1371
DB 1395 CAGAGCTTGAAGAGCAACCCGTTTGAATGGAAAGACTCGGTGCGAGAGCTCTC 1454
OY 1372 AATAGAGAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1431
DB 1455 AGTCAGAAAGACAGGAGAAACAGAAAGATTTGATGAGCTGAGCTCAGAAAGAAAGTCT 1514
OY 1432 GAGTTGATTAAGAGCTGATGATGACAAAGAGATCAGCTAGAGAGAAATTCAGAT 1491
DB 1515 CAGCTGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
OY 1492 ATCAGATGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
DB 1575 GTCCAAATCAGAAAGCAACAAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
OY 1552 CTGAAGATTTGTAATCACCCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
DB 1635 CTGAAATTTATGAAATCAACAACTTCAACAGAGAGAGAGAGAGAGAGAGAGAG 1694
OY 1612 GGAAGCTTATTTCCAGAGAAACAGATCTCAGTACAGAGAGAGAGAGAGAGAGAG 1671
DB 1695 ATCTATCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
OY 1672 AGTTGATAGAGAGCTGCTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAG 1731
DB 1755 AACACACTGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
OY 1732 CGGAGAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
DB 1815 TGCCAAAGACTTAAGAAACATTAAGATGCTCTGAAAGAAAGAGAGAGAGAG 1874
OY 1792 GAGATGATGTTTCAACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
DB 1875 GAATGATGATTTAATCAATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
OY 1912 AGCTGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
DB 1995 AGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017

```

```

RESULT 2
US-09-907-969-72
; Sequence 72, Application US/0907969
; Publication No. US20030091580A1
; GENERAL INFORMATION:
; APPLICANT: Micham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Recter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Patrick
; APPLICANT: Hill, Paul
; APPLICANT: Albone, Earl
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C8
; CURRENT APPLICATION NUMBER: US/09/907,969
; NUMBER OF SEQ ID NOS: 596
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72

```

LENGTH: 2017
TYPE: DNA
ORGANISM: Homo sapiens
US-09-907-969-72

Query Match 13.1%; Score 477.4; DB 11; Length 2017;
Best Local Similarity 55.5%; Pred. No. 8,6e-127;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

QY 1 ATGGCTCAGTTTCCACACCTTTCCGCTGATGCTGATGCTGAGGCTAATCTGAGAG 60
DB 42 ATGGCTCAGTTTCCACACGATGAAATGAGGCGCAAAATGTGGGCTATTATCATCTGAA 101
QY 61 GAAAGGCCCAGAGATGACAGCAGTTCCTAGCCTGAGCCGATAGCGGATTTATTTACT 120
DB 102 GAAAGTACTAGATGATGATTAACAGTTGATTAATCTCAAACTTCAGAGAGTTACATTA 161
QY 121 GGTGATCAAGCGAGAACTTTTTCATCTGGGTTTACCTGAGCCTGTCTTAGCACA 180
DB 162 GGTGATCAAGCCGCTACTTTTTCCTACAGTCTGCGGCGCCGGTTTATAGCTGAA 221
QY 181 ATATGGCCCTTACCGACATGATTAAGATGAAAGATGATCAATGGAATTTTCTATA 240
DB 222 ATATGGCCCTTATCAGATCTGAAACAAGATGGAGATGACCAAGAGTTCTCTATA 281
QY 241 GCCATGAGCTTATCAAACTGAAGCTACAGAGATATCAGCTCCCTCCACACTTCCCTC 300
DB 282 GCTATGAATCTATCAAGTTAAAGTTGACAGGCGCAACGCTGCTGATCTCTCTCT 341
QY 301 GTCATGAACAGAACCACTG---GCTATTTCCAGTGCACAGCATTGTTGATAGAGG 357
DB 342 ATCATGAACCAACCCCTATGTTCTCTCCACTATCTCTGCTGTTTGGATGGAGAGC 401
QY 358 ATGTCTAGATG---CCACCACTGACAGCTGTTGCTCTGTCCTA-----400
DB 402 ATGCCAATCTGTCATTCATTCAGCCATTGCTCCAGTTGACCTTATGACCAACCTT 461
QY 401 -----TGGGCTCCATTCCAGTTGGAATGTCCTCAACCTCTAGTATCT 444
DB 462 TCTTCTGCTAATCTCAGAGACAGATATCTCTCCCTTAATGATGCTGCTCTCTAGT 521
QY 445 TCTGTCCTCCAGCAGAGCTGCTCCCTGCTTAAGCGGCTCTCCCTCATACAGCT 504
DB 522 TCTGTTAGTACATCTCATTAACAAATGAACTGCCAGTCTCATGACCTTATCAT 581
QY 505 CTGCTGCTTTCGCTGCTGACGCAATGAGCCAAAGTTCTTCTCAGAGATCT 564
DB 582 CCTTATCTCTTCAACATTTGCTCATGATCATCTTACAGCTGATGAGGAGATTT 641
QY 565 GGTCCAGGGTCACAATTTAACTAAGTTTACAGAGGCAATCATTCATGTCGCGACG 624
DB 642 GGTGCTGCTAGTATCCAGAAAGCCCACTCTGATTTAGATCTGATGCTCAACT 701
QY 625 GCCCTCCAG-----CAGCAGATGGGCTGTG 651
DB 702 TCTCTAATGCTTCCCTCAGGGAATCACTAAGACAGGAGCTCAAGTGGGAGTT 761
QY 652 CCTCATCTCAAGGCTGAATACAGGAGTTTCAACAGCCACGAACTATGAGT 711
DB 762 CCTCAGCTTCAAGATTTAAAGTATGCGCAAAATTTAATGTTAAGCAAAAGCATGAG 821
QY 712 GGACATTTAAAGGTTCCCGCAGCAAGATCTATCTCATGCAATGATTTACCCAGGCT 771
DB 822 GGATACCTCTCAGGTTTTCAGCTAAATATGCCCTCTTCAGTCAAACTCTCTCAACT 881
QY 772 CAGCTGCTTCAATGATGATCTTTCTGATTTGATGATCAAGATGAAACTCACTGAGAA 831
DB 882 CAGCTGCTTCAATGATGATCTTTCTGATTTGATGATGATGATGATGATGATGATG 941
QY 832 GAATTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 891
DB 942 GAATTTATCTGCGATGACCTCACTGATGATGCGCAAAAGCTGAGCAGGCACTACCACTG 1001

QY 892 GTCTGCTCCAGAAATATATCCCTCTCTCTCAGAAAGTTGCTCCGCGAGTGGAGT 951
DB 1002 AGCTTGCCTCCGAGGCTTGTCTCTCTCTTTCAGAGGAGGAGAAAGCAAG-----1050
QY 952 TCCGTATTAAGCTCTTCTCTGATGATCAGAGCTGCTGAGAGCCGTCTGATGAGAT 1011
DB 1051 -----TTGATCTGTATATGGAATCTGCTCTTCAATATCAGAAACACAAAGAA 1097
QY 1012 GAGCAGACCCAGAGAAAGAACTGCTGCTGATCTTTGAAATTAAGAGCGGGAATCTTC 1071
DB 1098 GA---AGAGCTTCAGAAAGAACTGCGAGTAACTTTGAGAGACAAAGGAAAGCCACTAT 1154
QY 1072 GAGCAGAGAGTGTGAGGCTGAGAGAGCCCGCAAGCCCTCTTGGAGAGCAGCCCAAA 1131
DB 1155 GAAAGAGAAACATGAGAGCTGAGAGAGCAAGCCAACTTTGATGAGAGAGCAGAGAG 1214
QY 1132 GAGCAGAGAGGCTTGGCTGAGTGAAGCGCGCCGAGCAGAGAGAGAGAGAGCGGAG 1191
DB 1215 GAGGCTGAAGCGAAAGCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
QY 1192 CAGAGACAGAGAGCAGCGGAGCTGAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAG 1251
DB 1275 CAGAGCAAGATGAG 1334
QY 1252 CTGAGCGGAG 1311
DB 1335 CTGAG 1394
QY 1312 CCGGAACTGGAAG 1371
DB 1395 CAG 1454
QY 1372 AATCAG 1431
DB 1455 AGTCAAG 1514
QY 1432 GAGTTGATGAGAGCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1491
DB 1515 CACCTGAG 1574
QY 1492 ATCAGGTGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
DB 1575 GTCCAAATCAGAAAG 1634
QY 1552 CTAGAGATGCTGAATCAACCACTTACAGCAGAGCTGAGAGAGAGAGAGAGAGAG 1611
DB 1635 CTGAAATTTATGAAATCAAACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694
QY 1612 GGAAGCTTATTCAG 1671
DB 1695 ATCTATCTGCTGAG 1754
QY 1672 AGTTGATAGAGAGCTGCTCTTACCTCAAAAGAGCTTGAAGAGAGAGAGAGAGAG 1731
DB 1755 AACACAGCTGATTCAGAGAGATGATTTACTTCAATTAAGTCAATCAAGAGAGAGAG 1814
QY 1732 CCGCAGAGAGCTCCGAG 1791
DB 1815 TCCCAAGAGCTTAAAG 1874
QY 1792 GAGATTTGATTTTCAACAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1851
DB 1875 GAAATGAGATTTATTAACATGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1934
QY 1852 CTCAG 1911
DB 1935 TTAGGCTTGAAGCACTTCAATTAATCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1994
QY 1912 AGCTGAGAGTTGAG 1934
DB 1995 AGATTAGAGCAAAAAAAG 2017

RESULT 3
US-09-827-271-72
; Sequence 72, Application US/09827271
; Publication No. US20030165504A1
; GENERAL INFORMATION:
; APPLICANT: Recter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827.271
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO: 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-827-271-72

Query Match 13.1%; Score 477.4; DB 12; Length 2017;
Best Local Similarity 55.5%; Pred. No. 8.6e-127;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

QY 1 ATGGCTCAGTTCCACACCTTTCGGTGTAGCTGATGTCGGGCCATTAATCTGTGAG 60
DB 42 ATGGCTCAGTTCCACACCTTTCGGTGTAGCTGATGTCGGGCCATTAATCTGTGAG 101
QY 61 GAAAGGGCCAAAGCATGACGACGATTCCTTAGCTGAAGCCGATAGCGGGATTATTA 120
DB 102 GAAAGGGCCAAAGCATGATGATTAAGTTGATTAACCTTAAGCTTGAAGGATTAC 161
QY 121 GGTATTCAGAGGAGAACTTTTTCCTCAATCTGGCTTACTCTGACCTGTCTTAC 180
DB 162 GGTATTCAGAGGAGAACTTTTTCCTCAATCTGGCTTACTCTGACCTGTCTTAC 221
QY 181 ATATGGGCGCTACGCGACATGATTAAGTGAAGATGATGATCAAGTGAATTTTCC 240
DB 222 ATATGGGCGCTTATCAATCTGATGATTAAGTGAAGATGATGATCAAGTGAATTTT 281
QY 241 GCCATGAAGCTTATCAATCTGATGATTAAGTGAAGATGATGATCAAGTGAATTTT 300
DB 282 GCTATGAAGCTTATCAATCTGATGATTAAGTGAAGATGATGATCAAGTGAATTTT 341
QY 301 GTCATGAAGCAAGCAACGCTG---GCTATTTCCATGCAACGACATTTGGTATG 357
DB 342 ATCATGAAGCAAGCAACGCTG---GCTATTTCCATGCAACGACATTTGGTATG 401
QY 358 ATGTGCTAGCATG---CCACACATCAAGCTGTTGCTCCTGTGCA--- 400
DB 402 ATGTGCTAGCATG---CCACACATCAAGCTGTTGCTCCTGTGCA--- 461
QY 401 -----TGGGCTCATTCCAGTTGTTGGAATGTCACACCTTAGTATCT 444
DB 462 TCTTGTGCTAGTTCCAGGACGATGATTCCTCCCTAATGATGCTGCTCCCTAGT 521
QY 445 TCTGTCCCTCCAGCAGCTGCTCCCTGGCTTAAAGGGGCTCTCTCCGCTATAC 504
DB 522 TCTGTGTAGTACATCTCTCAATTAACCAATGGAATGACATCTCAAGCTTATCA 581
QY 505 CTGCGCTGGCTTGGCGATCTGCGACGACATGGCCAAAGAGTTCTTCCCTTAC 564
DB 582 CTTATTTCTTCTTCAACATGCTGCTATGATCATCTTACAGCTGATGATGAGG 641
QY 565 GGTTCAGAGGTCAATTAACATTAAGTTACAGAAAGCAATCAATCTGATGCTCC 624
DB 642 GGTGTGTAGTATCAGAAAGCCAGTCTGATGATTAAGTTAGATTTAGTAC 701
QY 625 GCCCTCCAG-----CAGCAATGGGCTGTG 651
DB 702 TCTCTCACTGCTCCCTCTCAGGAACTACCTTAAGACAGGAGCTCAGAGTGG 761
QY 652 CCTCATGATCAAGGCTGAATTAAGGACGATTAATCAACAGCAAGCAAAATCT 711

DB 762 CCTCAGCTTCAAGATTAAGTATCGGCAAAATTTATATGTCATGACAAAGCATAG 821
QY 712 GGAACCTTAACAGTCCCCAGGCAAGATTAATTCATGCAATCAAGTTATCCAG 771
DB 822 GGAATCCTTCAAGGTTTCAAGCTAGAAATGCCCTTCTTCAAGTAAATCTCTCA 881
QY 772 CAGTGTCTCAATATGGAATCTTTGACATGATGATGATGAAATCACTGACGAA 831
DB 882 CAGTGTCTCAATATGGAATCTTTGACATGATGATGATGATGAAATCACTGAC 941
QY 832 GAATTTATCTAGCTATGACCTTAATGATGTCCTGATGTCGTCAGCCATCCGCC 891
DB 942 GAATTTATCTAGCTATGACCTTAATGATGTCCTGATGTCGTCAGCCATCCGCC 1001
QY 892 GTCTGCTCCCAAGTATCACTCTCTCTTCTTCAAGAAAGTTCCTCCGCAATG 951
DB 1002 ACGTTGCTCCCGAGCTTGTCTCCATCTTTCAGAGGGGGAAGCAAG----- 1050
QY 952 TCCGCTAATAGCTCTTCTGCTGTGATCAAGGCTGCTGAGAGCGCTGTCAGAG 1011
DB 1051 -----TTGATCTGTAAATGAACTTGCCTTCATATCAAGAAACCAAGAA 1097
QY 1012 GAGCAGACGCGCAAGAAATCTGCTGTGATTAATGAAATGAAGCGGAGAACTTC 1071
DB 1098 GA---AGAGCTCAGAAAGAACTGCCAGTTACTTTTGAAGCAAAAGGAAAGCA 1154
QY 1072 GAGCAGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATG 1131
DB 1155 GAAAGGAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATG 1214
QY 1132 GAGCAGACGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
DB 1215 GAGCTGAAGCAAGCGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1274
QY 1192 CAGAGAGAGAGCGCAAGCGGCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1251
DB 1275 CAAGAGCAAGATGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1334
QY 1252 CTGAGCGGCAAG 1311
DB 1335 CTGAG 1394
QY 1312 CGGAACTGGAAG 1371
DB 1395 CAG 1454
QY 1372 AATCAG 1431
DB 1455 AGTCAAG 1514
QY 1432 GAGTTGATGATGAG 1491
DB 1515 CACCTGGAAG 1574
QY 1492 ATCAGGTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
DB 1575 GTCAAAATCAAG 1634
QY 1552 CTAAGAAATGCTGAATCAACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1611
DB 1635 CTGGAATTAAG 1694
QY 1612 GAAAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1671
DB 1695 ATCTATCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1754
QY 1672 AGTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
DB 1755 AACAACCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1814
QY 1732 CGGAGAGAGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791

Db 1815 TGCCAAAGACTTAAAGAACATTAAGTCTCTGTAAGAAAAAAGCATCTAAGCTTCA 1874
Qy 1792 GAGATTGATGTTTCAACCAAGCGTGAAGAACTGAGAGATCATAGCAAAACAGCA 1851
Db 1875 GAAATGATTTCAATTACATAGCTGAAGAACTCAGAGAACTATATACAGCAG 1934
Qy 1852 CTCGAAGACAGAGTCCCTGAGGACGACCTGAGAGAAAGAGAGAGAGAG 1911
Db 1935 TTAGCCCTTGAACAATTCAAAAGTGAAGAAATGAGAGAAATGAGAGAA 1994
Qy 1912 AGCTGAGATTGAGAGCAAAA 1934
Db 1995 AGATTAGCAAAAAA 2017

RESULT 4
US-10-198-053-72
; Sequence 72, Application US/10198053
; Publication No. US20030124140A1
; GENERAL INFORMATION:
; APPLICANT: Bangor, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-198-053-72

Query Match 13.1%; Score 477.4; DB 14; Length 2017;
Best Local Similarity 55.5%; Pred. No. 8.6e-127;
Matches 1111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;

Qy 1 ATGGCTCAGTTTCCACACCTTTCGCTGAGCTGATCTGAGCCATTAAGTGAAG 60
Db 42 ATGGCTCAGTTTCCACACGATGATGAGAGGCCAAATATGAGCTTATCATCTGA 101
Qy 61 GAAAGGCGCAAGCATGACGAGAGTCTTACCTGAGCCGATGAGCGGATTTACT 120
Db 102 GAACTACTAAACATGATTAACAGTTGATTAACCTCAAACTTCAGAGATTAACA 161
Qy 121 GGTGATCAAGGAGAACTTTTTCATCTGAGTTACCTCAGCTGTCTTAGACAA 180
Db 152 GGTGATCAAGCCCGTACTTTTCTTCTACAGTGTGCGGCGCCGCTTTAGCTGA 221
Qy 181 ATATGCGCGCTACCGACATGATTAAGATGAGATGATCAAGTGAATTTTCATA 240
Db 222 ATATGCGCGCTTATCATGATGATTAAGATGAGATGATCAAGATTTCTTATA 281
Qy 241 GCGATGAAGCTTATCAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 300
Db 282 GCTATGAAGCTCATCAAGTTAAAGTTGACAGGCCAAACAGCTCTGTATCTCTCCCT 341
Qy 301 GTCATGAAGACAGAACCACTG---GCTATTTCCAGTGCACAGCAATTTGATAGAAG 357
Db 342 ATCAATGAAGAACCCCTATGTTCTCTCACTAATCTGCTGTTTGGAGTGAAGC 401
Qy 358 ATTTGCTAGCATG-----CCACACTCAGAGCTGTGCTCTGTGCGAA----- 400
Db 402 ATGGCCAAATCTGTTCATTCATGACCATTTGCTTCAAGTTGACATAGCAACCCCTG 461
Qy 401 -----TGGGCTTCATTCAGTTTGGATGTCTTCACCTTAATCT 444
Db 462 TCTTCTACTACTTCAAGGACAGATATCTCTCCCTAATGATGCTGCTCCTAGTGCCT 521

Qy 445 TCTGTCTCCAGCAGCAGCTGCTCCCTGCTTAACGGGGCTCTCCGCTATACAGCT 504
Db 522 TCTGTATGATACATCTCTATTAACAAATGAGATGCTGCTATCATCTTATCATCT 581
Qy 505 CTGCTGCGTTTGGCATCTGACGACCAATGCGCCAAAGAGTTCTTCTTACAGCATCT 564
Db 582 CTTATTTCTTCTTCAACATGCTGCTATCATCTTATCATCTGATGAGAGATTT 641
Qy 555 GGTCCAGGATCAATTAACATTAAGTTACAGAGCAATCACTTGTATGCTGCCAGC 624
Db 642 GGTGTGTAGTATTCAGAGGCCCAATCTGATTTAGTTAGATGATGATGCTCACT 701
Qy 625 GCCCTCCAG-----CAGCAAAATGGGCTGTG 651
Db 702 TCTTAATCTGCTCTCTCTCAGGAACTCAGTAAAGAGAGAGCTCAGAGTGGCATTT 761
Qy 652 CTTCAATCATCAAGGCTGAAATACAGGAGTTTCAACAGCCAGCAAGCAAACTATGAGT 711
Db 762 CTTCAAGCTTCAAGATTAAGTATGCGCAAAATTAATGATGATGACAAAGGATGAGC 821
Qy 712 GACACTTAACAGGTTCCCGAGCAAGAACTATCTCATGCAATCAAGTTTACCCAGCT 771
Db 822 GATACCTCTCAGGTTTCAAGCTAAGATGCTTCTTCAAGTCAAACTCTCTCAACT 881
Qy 772 CAGCTGCTTCAATATGATCTTCTGATGATGAGAGGCTGCTGAGAGCCGCTCAGAGAT 831
Db 882 CAGCTAATCTATTTGATGATCTGCTGATCATGATGATGATGATGATGATGATGAT 941
Qy 832 GAATTTATCTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 891
Db 942 GAATTTATCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001
Qy 892 GTCTGCTTCAAGATTAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 951
Db 1002 AGGTGCTCTCCGAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1050
Qy 952 TCCGTATAGCT 1011
Db 1051 -----TTGATTTCTTATGAACTCTGCTTCTATCAATTAACAAAGCAAGAA 1097
Qy 1012 GAGCAGCAGCAGAGAGAACTGCTGTGACATTTGAAGATTAAGAACGGAGAACTTC 1071
Db 1098 GA---AGAGCTCAGAAAGAACTGCACTTCTTTTGAAGCAAAAGCAAGCACTAT 1154
Qy 1072 GAGCAGCAGCTGTGAGCTGAGAGAGCGCGCAAGCGCTCTTGAAGCAGCGCAAA 1131
Db 1155 GAAAGAGAGAACTATGATCTGAGAGAGCGAGCCAAAGTTGATGAGCAGCAGAGAG 1214
Qy 1132 GAGCAGAGCGGTTGGCTCAGCTGAGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAG 1191
Db 1215 GAGGCTGAACGAAAGCCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1274
Qy 1192 CAGGAGCAGAGAGCGCAAGCTGAGCTGAGAGAGCAGCTGAGAGAGAGAGAGAGAG 1251
Db 1275 CAGAGCAAGATGAG 1334
Qy 1252 CTGAGCGGCAAGCAG 1311
Db 1335 CTGAG 1394
Qy 1312 CGGGAACGGAAG 1371
Db 1395 CAGAGCTTGAAG 1454
Qy 1372 AATCAG 1431
Db 1455 AGTCAAG 1514
Qy 1432 GAGTTGATTAAG 1491
Db 1515 CAGCTGAG 1574
Qy 1492 ATCAGGTTTCAAGTGAAG 1551

Db 1575 GTCCAAATCAGAAAGCAAAAGCTGAGTAAAGTTTGGATTAACAGTGTGAC 1634
Qy 1552 CTAAAGATTGCTGAATACCACTTACAGCAGCAGTTGCGAATCTCAGCAATGCTT 1611
Db 1635 CTGGAATTTATGGAATCAACCACTTCAACAGAGCTTAAAGAAATATCAAAATAGCTT 1694
Qy 1612 GGAAGACTTATTCAGAGAAACAGATCTCAGTGCAGTAAACCAAGTCAGAGAAC 1671
Db 1695 ATCTATCTGCTGCTCAGAGAGCAGTATTAACGAAAGAAATTAATAATGACAGCTCAGT 1754
Qy 1672 AGTTGATAGAGAGCTGCTCTTTAACCCTCAAAAGAGCTTGAAGCAAGAGAGTGGCC 1731
Db 1755 AACACAGCTGATTCAGGATCAGTTTACTTATTAAGTAAGTATGAGAAAGAAATTA 1814
Qy 1732 CGGAGCAGCTCCGGAGCAGCTGAGCAGAGTGAAGAGAGACAGTCAAGCTGAG 1791
Db 1815 TGCCAAAGACTTAAAGAACTTATGATGCTCTTGAAGAAAGAACTGATCTTCA 1874
Qy 1792 GAGATGATGTTTCAACACAGCTGAGAGAACTGAGAGATCATAGCAACAGCA 1851
Db 1875 GAAATGATTCATTTAAACAATCAGCTGAAGAGAACTCAAGAAAGCTATTAATACAGCAG 1934
Qy 1852 CTCAGAGCAGAGGCTCCCTGAGCAGCAGCTGAAGCAAGAAAGCAGAGAGAG 1911
Db 1935 TTAGGCTTGAACAACCTTATTAATCAAGCTGACAAATTTGAAGAAATGAGAAAGAAA 1994
Qy 1912 AGCCTGAGTTAGAGAGCAAA 1934
Db 1995 AGATTAGAGCAAAAAA 2017

RESULT 5

US-09-764-881-55
; Sequence 55, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT207
; CURRENT APPLICATION NUMBER: US/09/764,881
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (481)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (536)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (556)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (562)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-881-55

Query Match 9.8%; Score 356; DB 11; Length 568;
Best Local Similarity 86.2%; Pred. No. 5, 2e-92;
Matches 426; Conservative 0; Mismatches 64; Indels 4; Gaps 3;

Qy 1 ATGGCTCAGTTTCCACACCTTTGCTGCTAGCCTGAGTGTGGGCAATATCTGTGAG 60
Db 78 ATGGCTCAGTTTCCACACCTTTGCTGCTAGCCTGAGTGTGGGCAATATCTGTGAG 137
Qy 61 GAAAGGGCCAAAGATGACCAAGAGTCTTACCTGAAGCCGATAGCGGATTTTACT 120

Db 138 GAAAGAGCGA--CATGATCAGCAGTTCATAGTTTAAAGCCAAATCTGATTCATTA 195
Qy 121 GGTGATACACAGAGAACTTTTTCATCTGGGTACTAGCTGCTTAGACAA 180
Db 196 GGTGATACACAGAACTTTTTCATCTGGGTACTAGCTGCTTAGACAA 255
Qy 181 ATATGGCGCTAGCGGACATGAATTAAGATGAAGATGATGAATTTTCCATA 240
Db 256 ATATGGCGCTAGCGGACATGAATTAAGATGAAGATGATGAATTTTCCATA 315
Qy 241 GCGATGAGCTTATCAAACTGAAGCTAAGATATCAAGTCTCCCTCCACACTTCCCT 300
Db 316 GCTATGAACTTATCAAACTGAAGCTAAGATATCAAGTCTCCCTCCACACTTCCCT 375
Qy 301 GCTATGAACTTATCAAACTGAAGCTTATTCAGTGCACACCACTTGTATAGAGAGAT 360
Db 376 GCTATGAACTTATCAAACTGAAGCTTATTCAGTGCACACCACTTGTATAGAGAGAT 435
Qy 361 GCTATGAGCTTATCAAACTGAAGCTTATTCAGTGCACACCACTTGTATAGAGAGAT 420
Db 436 GCGAGCAGCAGCAGCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 495
Qy 421 -GGATGCTCCACCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 479
Db 496 GGGATGCTCCACCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 554
Qy 480 CGGAGCTCTCCG 493
Db 555 AMGGGTTNCCCTG 568

RESULT 6

US-09-879-957-193
; Sequence 193, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOPFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOWLES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrick, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE

QY	3149	TTAAAGATTCAAGAGGCTCTGGAACTGCTGGGAAAAACAGGGAGTTTACGAAAAAAACCTG	3208
Db	1437	CMAAGATTCMAAGAGGTTTGGGAGCTTAGCAAGCTCTGGAGCATCAATAAANAACCTG	1498
QY	3209	AAATTGCCAGATTATTCCTTCTTAAGCTGCTACTGTCCTCCGAACACTCACTTGAGCTC	3268
Db	1497	AGATTGCTCAGGTAACCTCAGCATATGTGTGCTTCTGGTTCTGAAACAACTTAGCCTTGAC	1556
QY	3289	CTGGGCACTGATCTGATCCGGAAAAAGAACCCGAGTGGATGGTGGGAAGAGAACTCG	3328
Db	1557	CAGGACATTAATATTAATTTCTAAAGAAAAATACAGTGGGTGGTGGCAAGAGAGTTAC	1616
QY	3329	AAGCTCAGGAGAAAAAGCCAGATAGGGTGGTTTCAGCAAAATTATGTCAAATCTTCAA	3388
Db	1617	AGGCACAGAGAAAAAGCAGACAGAAAGATGGTTTCTGCAGATCATGTTAAACTTTTGG	1676
QY	3389	GCCCCGGAACAAGCAAAATTCACCCCACTAGACTACCAAGACCGCACTGACGACGAG	3448
Db	1677	GTTCCAGTAGTGAAAGAGCCACACCTGCTTTTCATC-----CTG	1715
QY	3449	TGTGTCAGGTGATCGGAGTGTACGATTAACACCGCCAGAGACGATGACAACTGACCTTCA	3508
Db	1716	TATGTCAGGTGATTCCTATGTATGATGACTATGCGCAAAATTAATGAAGATGAGCTCAATTTCT	1775
QY	3509	GCAAAGGCAATCATCTCAACGTCTTCCAAACAGAGAGAACCCGGACTGTGTGAAAAAGAGAG	3566
Db	1776	CCAAAGGCAACTCATTAATGTATTAAACAAAGATGATCTGATTTGGTGGCAAGAGAGA	1835
QY	3569	TCAGTGGGCAAGTTGGGCTCTTCCCATCAATTATGTAAAGCTGACCAAGACATGAGAC	3628
Db	1836	TCAAGGGGTGATGCTGTCTCTTTCTTAAACTAGCTTAAGATGACGACAGCTCAGATC	1895
QY	3629	CAAGCCAGCAATG	3641
Db	1896	CAAGTCAACAGTG	1908

RESULT 8

```

Sequence 31258, Application US/09918995
Publication No. US20030073622A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 31258
LENGTH: 503
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(503)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-31258

```

Query Match	8.9%	Score 322.6;	DB 11;	Length 503;
Best Local Similarity	81.6%	Pred. No. 2.2e-82;		
Matches 386;	Conservative 0;	Mismatches 84;	Indels 3;	Gaps 1

QY	823	CTGGAGAGAAATTTATCTAGTAGACACCTAATGATGTTGCCANGTCTGTACGCC	882
Db	30	AATTGAGAGAAATTTATCTCTGCAATGCAACTATTGATGTACCTATGTCTGGCAACCA	89
QY	883	CTGCCGCCCGTCTCTGCTCAAGATACATCCTCTTCTCTTCAGAGAGTTGGCTCGGC	942
Db	90	CTGCCACCTGTCTCTGCTCTCAAGATACATTCACCTTCTTTTGAAGAGTTCATCTGGC	149

Qy	943	AGTGGGATGTCGGCATTAAGCTCTTCTTGATGATCAGAGGCGTCCGTAGAGACCCCTCG	1002
Db	150	AGGTGATATCTGTCATTAAGCTCAACATCTGTAGATCAGAGCTTACCAAGAGAACCACTT	209
Qy	1003	TCAGAGATGAGCAGCAGC---CAGAGAGCAATGTGCTGTGACATTTGAATTAAGAG	1058
Db	210	TTAGAAATGACACACAACTTTGAAAAAGAAATTACTGTGAACGTTTGAAGATTAAGAG	269
Qy	1060	CGGAGAACTTTGAGCGAGGCAGTGTGAGCTGAGAAACGCGCAAGCGCTTTTGAG	1119
Db	270	CGGAGAACTTTGAACTGTGCAACTGTGAACCTGGAACGAAAGCAAGCTCTCTGGAA	329
Qy	1120	CAGCAGCGCAAAAGCAGAGCGCGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAA	1179
Db	330	CAGCAGCGCAAGAGCAGAGCGCGCTTGCCCACTGTGAGCGCGCGAGCAGAGAGAGAG	389
Qy	1180	GAGCGGAGAGCGCCAGAGAGCAGAGAGGCCAAGCGGAGCTGAGAGCTGGAAGACAGCTGGAG	1239
Db	390	GAGGTTAGCGCCAGAGAGCAAGACCGATTAAGACAACTGGAACCTGAGATCAACTGGGA	449
Qy	1240	AAGCAGCGGAGCTGAGAGCGGCAAGAGAGAGAGAGAGAGAGATGCA	1292
Db	450	AAGCAGCGGAGCTGAGAGCGGCAAGAGAGAGAGAGAGAGAGATATTGCA	502

RESULT 5

US-09-879-957-39
Sequence 39, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFEMAN, No. US20020034755A1b

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME

NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:

**ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas**

CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMMITTED: IBM PC format ibj

COMPIER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0 Version #1.30

SOFTWARE: FACELIN RELEASE #1.0,
CURRENT APPLICATION DATA:
APPLICATION NUMBER: IIS/09/879.957

ALLOCATION NUMBER: 00/00/00/00
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:

NAME: Mibrook, S. Leslie
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 39:

SEQUENCE CHARACTERISTICS:
LENGTH: 747 bases

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLGY: linear

TOPOLOGI: LINEAR

MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-879-957-39

Query Match 7.1%; Score 258; DB 9; Length 747;
Best Local Similarity 63.9%; Pred. No. 1.3e-63;
Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

```

QY 2989 GGAGAGAGTTTATTCATGTCACATGACAGAGTTTCGAGCAAGATTAACTT 3048
DB 115 GGAGAGAGTTTATTCATGTCACATGTCAGAGTTTCGAGCAAGATTAACTT 174
QY 3049 CAGCAAGAGGAGTGTGATTGTTACCAAGAAAGATGTCAGTGTGACGAGGACGTG 3108
DB 175 ACAGAAAGTGAAGAAATATTTGGTGAACCAAGAAAGATGTCAGTGTGACGAGGACGTG 234
QY 3109 GCGACAGATCCGAGTCTTCCCTTCTAACTATGAGGCTTAAAGATTGAGGGCTCT 3168
DB 235 GGAGATGAGAGTGAATTTTTCATCAAACTATGTCAAACTAAGATCAAGAGATT 294
QY 3169 GGAACCTGCTGGGAAAACAGGAGTTAGGAAAAAACTGAAATTCGACGATTATGCT 3228
DB 295 GGGAGTGTCTAGCAAGTCTGAGCATCAATATAAAACTGAGATTGCTCAAGTAACTTCA 354
QY 3229 TCTTACGCTGTACTGTGTCCTCCGAACTACCTGCTCTGAGGAGCTGATTCTGATC 3288
DB 355 GATATATGTTGCTTCTGTGTTCTGAACAATTAGCTTGCACAGACAGATTAAATTAAT 414
QY 3289 CCGAAAAAGAACCCAGGTGATGTGTGGAAAGAACTGCAAGCTGAGGAAAAAGCCG 3348
DB 415 CTAAGAAAAAATACAGATGAGTGTGTGGCAAGAGATTACAGCCAGAGAAAAAGCGA 474
QY 3349 CAGATAGGAGTGTTCACCAAAATTATGTCAAACTTCAAGCCCGGAAACAGCAAAATC 3408
DB 475 CAGAAAGAGTGTTCCTCCAGTCAATGTTAACTTTGGTCCAAAGCAGTGAAGAGCC 534
QY 3409 ACCCCAACTGAGTACCCCAAGACCCGAGTGCAGCCAGAGTGCAGGTGATCGGAGT 3468
DB 535 AACCTGCTTTCATCTCT-----GTAATGTCAGTGTGATGCTAAG 573
QY 3469 TACGATTACCGCCCAAGACAGTACGAACTAGCCTTACGCAAAAGCCAGATCATCAAC 3528
DB 574 TATGACTATGACAGAAATATATGAAATGAGTCACTGATTTCCCAAGGAGCAACTAATAT 633
QY 3529 GTCCCTCAAGAGAGACCCCGAGTGTGTGGAAAGAAAGTCAAGTGTGGCAAGTGGCTC 3588
DB 634 GTTATGAAACAAAGATATCTCTGATGTGTGGCAAGAGATCAACCGGGGTGACTGTGCTC 693
QY 3589 TTCCCATCCATATATATTAAGCTGACCAAGATGAGACCCCGCAGCAATGA 3642
DB 694 TTTCCTTCAAACTACTGTTAAGATGACAGACACTCAGATCCAGTCAACAGTGA 747

```

RESULT 10
US-09-864-761-17127
Sequence 17127, Application US/0964761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR APPLICATION NUMBER: 2001-05-23
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03

```

PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 17127
LENGTH: 270
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000193.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EST HUMAN HIT: AA773263.1, EVALUE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
OTHER INFORMATION: NT HIT: g11526214, EVALUE 0.00e+00
US-09-864-761-17127

```

Query Match 5.5%; Score 198.8; DB 9; Length 270;
Best Local Similarity 87.2%; Pred. No. 9.8e-47;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 2299 GTGATGAAGCCAGACTGAGAGCCAGAGTGGCTTGAGAGAGCTGAAGGAAAGC 2358
DB 1 GTGATGAAGCCAGACTGAGAGCCAGAGTGGCTTGAGAGAGCTGAAGGAAAGC 60
QY 2359 GGATGTTCCCTCGCAACTATGAGAAAAAGATTCCAGAAAAATGAGTTCCCACTCAGCC 2418
DB 61 GGGTGTTCCTCGCAACTATGAGAAAAATCCAGAAAAATGAGTTCCCGCTCAGTG 120
QY 2419 AAACAGTGAACGATCTGACATCTGCTCCCTGCCCCCAACTGCTGTGAGAGACCT 2478
DB 121 AAACAGTGAACGATCTGACATCTGCTCCCTGCCCCCAACTGCTGTGAGAGACCT 180
QY 2479 GCTCTTTGCAAGTACCTCTTTGAGCTTTCACAAACCCCACTGAGGAGAGCTTC 2538
DB 181 GCCCTTTGAGAGTAACTCTTCAAGAGCCCTTCAAGACCCCTTAATTAATGAGGAGCTTC 240

```

QY 2539 AGTCCACGT 2548
|||
Db 241 AGTCCACGT 250

RESULT 11
US-09-864-761-26948
Sequence 26948, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26948
LENGTH: 286
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
OTHER INFORMATION: NT HIT: AF114488.1, EVALUATE 0.00e+00

OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATE 2.00e-45
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUATE 1.00e-112
US-09-864-761-26948
Query Match 5.5%; Score 198.8; DB 9; Length 286;
Best Local Similarity 87.2%; Pred. No. 1e-46;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 2299 GTGATGAAAGCCGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAACG 2358
|||
Db 1 GTGATGAAAGCCGAACTGGAGAACCCGGCTGTGAGAGATTAAAGGAAAGCA 60
QY 2359 GGATGTTCCCTGCAACTATGACAAAGATTGCAAAATGATGCTCCACTCCAGCC 2418
|||
Db 61 GCGTGTCTTCTGCAAACTATGACAAAGATTGCAAAATGATGCTCCACTCCAGTG 120
QY 2419 AAACGATGACCGATGCACTGACATGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478
|||
Db 121 AAACGATGACCGATGCACTGACATGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180
QY 2479 GCTCTTTGCAAGTACCTCTTGGAGCCCTCCCAACCCCAACACTGCGGAGACTTC 2538
|||
Db 181 GCCCTTTGGCAGTAACTCTTTCAGAGCCCTCCAGACCCCTAAATGATGCGGAGACTTC 240
QY 2539 AGTCCACGT 2548
|||
Db 241 AGTCCACGT 250
RESULT 12
US-09-864-761-30453
Sequence 30453, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecolica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30453
LENGTH: 297
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AA73263.1, EVALUE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-864-761-30453

```

```

Query Match
Best Local Similarity 5.5%; Score 198.8; DB 9; Length 297;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAGAAGCAGAGTGGAGAGGAGGAGTGGAGAGAGTGAAGAGAGAGG 2358
DB 1 GTGATGAAGAAGCAGAGTGGAGAGGAGGAGTGGAGAGAGTGAAGAGAGAGG 60
QY 2359 GGATGTTCTCTGCAAACTATGAGAGAGAGTTCAGAAATGAGTTCCACTCCAGCC 2418
DB 61 GGATGTTCTCTGCAAACTATGAGAGAGAGTTCAGAAATGAGTTCCACTCCAGG 120
QY 2419 AAACGAGTACCGATGATGATGATGATGATGATGATGATGATGATGATGATG 2478
DB 121 AAACGAGTACCGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 2479 GCTCTCTCTGCAAGTCTCTCTGAGCCCTCCAGAACCCCAACAACCTGGAGACTTC 2538
DB 181 GCTCTCTCTGCAAGTCTCTCTGAGCCCTCCAGAACCCCAACAACCTGGAGACTTC 240
QY 2539 AGTTCACGT 2548
DB 241 AGCTCCAGCT 250

```

```

RESULT 13
US-09-864-761-17146
Sequence 17146, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6

```

```

PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17146
LENGTH: 301
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUE 3.50e-01
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 1.00e-122
OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUE 4.00e-62
US-09-864-761-17146

```

```

Query Match
Best Local Similarity 5.5%; Score 198.6; DB 9; Length 301;
Matches 228; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1040 TGACATTGGAATTAAGAGGAGGAGAACTTGAAGCGGAGTGTGAGTGAAGAAC 1099
DB 1 TAACTTTGAAGATTAAGAGGAGGAGAACTTTGAAGCGGAGCACTGGAACCTGAGAAAC 60
QY 1100 GCCGCAAGCGCTCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1159
DB 61 GAGGCAAGCTCTCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 1160 GCGCGAGCAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1219
DB 121 GGGCGAGCAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
QY 1220 AGCTGAGAGAGAGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAG 1279
DB 181 AACTGAGAGAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAGAG 240
QY 1280 GAAAGAGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1316

```


Db 241 GGAAAGAAATTGAGAGCGAGAGGTAAGCAGCGAGA 277

```

RESULT 14
US-09-864-761-17644
; Sequence 17644, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17644
LENGTH: 263
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST HUMAN HIT: BB542917.1, EVALUATE 3.00e-62

```

; OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUE 2.80e-01
 ; OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 1.00e-122
 US-09-864-761-17644

Query Match	5.4%	Score 197.4;	DB 9;	Length 263;
Best Local Similarity	84.4%	Pred. No. 2.5e-46;		
Matches 222;	Conservative	0;	Mismatches 41;	Indels 0;
				Gaps 0;

[illegible]

```

RESULT 15
US-09-864-761-10314
/ Sequence 10314, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
/ FIDE REFERENCE: Aecmca-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 4917
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 10314
LENGTH: 480
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
US-09-864-761-10314

Query Match 5.3%; Score 192; DB 9; Length 480;
Best Local Similarity 81.6%; Pred. No. 1.3e-44;
Matches 222; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2265 TGAGATCACCATCCAGCCAGAGATATAGTCATGTGATGAAGCCAGACTGAGAGCC 2324
DB 209 TTAGTTCTACTTAAGCTGTGACTTTTCCACAGGTGATGAAGCCAACTGAGAGCC 268
QY 2325 AGGATGGCTTGGAGAGAGCTGAAGGAAGAGCGATGTTCCCTGCAACTATGACAG 2384
DB 269 CGGCTGCTTGGAGAGAAATTAAGGAAGACAGAGGTGTTCCCTGCAACTATGACAG 328
QY 2385 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2444
DB 329 GAAATTCAGAAATGAGGTTCCCGCTCCAGTAAACAGTATGATTCACATCTGC 388
QY 2445 CCCTGCCCCCAACTGGCTCTGCGTGAGACCCCTGCTCTTGGCAGTGACCTTCTGA 2504
DB 389 CCCTGCCCCCAACTGGCTCTGCGTGAGACCCCGCCCTTGGCAGTAACTCTTTCAG 448
QY 2505 GCCCTCCACAAACCCCAACAACTGGGCAACT 2536
DB 449 GCCCTCCACAAACCCCAACAACTGGGCGCACT 480

Search completed: December 5, 2003, 06:15:46
Job time : 777.554 sec

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2003, 15:21:40 / Search time 1203.2 Seconds

(without alignments)
17313.032 Million cell updates/sec

Title: US-09-674-237A-1

Perfect score: 5084

Sequence: 1 cgcacgagagagagtggtgag.....gagaatcgatatacaagcct 5084

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: GenBank:1:
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_in:*
32: em_hcg_ocher:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vit:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_ocher:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3721.4	73.2	3723	10	AF132478 Mus muscu
2	3639.4	71.6	5145	10	AF132481 Mus muscu
3	3529.4	69.4	5381	9	AF114488 Homo sapi
4	3517.2	69.2	4025	10	AF127798 Rattus no
5	3440.2	67.7	5287	9	AF064243 Homo sapi
6	3103.6	61.0	3812	10	AF132672 Rattus no
7	2975	58.5	6439	9	AF114487 Homo sapi
8	2898.4	57.0	7247	9	AF064244 Homo sapi
9	2047.2	40.3	4103	5	AF032118 Xenopus l
10	1610	31.7	2131	6	BD127168 Primer fo
11	1610	31.7	2131	9	AK074554 Homo sapi
12	1404.2	27.6	2131	6	BD158570 Primer fo
13	1404.2	27.6	2131	9	AK027846 Homo sapi
14	1252.6	24.6	3241	9	HS061166 Human SH3 d
15	1249.2	24.6	1676	6	BD127640 Primer fo
16	1249.2	24.6	1676	9	AK075290 Homo sapi
17	1110	21.8	1133	10	AF169621 Mus muscu
18	1039.4	20.4	212305	2	AC134837 Mus muscu
19	1039.4	20.4	230097	2	AC126053 Mus muscu
20	981.8	19.3	1299	9	BC039036 Homo sapi
21	883	17.4	1996	9	AF180522 Homo sapi
22	813.6	16.0	78190	9	AC106760 Homo sapi
23	813.6	16.0	179264	2	AC023112 Homo sapi
24	731.4	14.4	5938	9	AB033082 Homo sapi
25	726.4	14.3	5828	6	BD167848 Method fo
26	726.4	14.3	5828	9	AF248540 Homo sapi
27	724.2	14.2	4557	9	AF182199 Homo sapi
28	701.2	13.8	278501	2	AC098552 Rattus no
29	701.2	13.8	279148	2	AC123507 Rattus no
30	679.6	13.4	135924	2	AC115529 Rattus no
31	679.6	13.3	4977	10	AF132479 Mus muscu
32	679.6	13.3	4977	6	AF132480 Mus muscu
33	673.8	13.0	4447	6	AX428899 Sequence
34	659	12.5	181782	2	AC116970 Rattus no
35	633.4	12.5	5804	10	AK122480 Mus muscu
36	617.4	12.1	877	6	BD124639 Primer fo
37	611.6	12.0	877	6	BD126208 Primer fo
38	611.6	12.0	877	6	AP000312 Homo sapi
39	556.2	10.9	12015	9	AP000193 Homo sapi
40	556.2	10.9	100000	9	AP000050 Homo sapi
41	556.2	10.9	114929	9	AP000117 Homo sapi
42	556.2	10.9	151516	9	AP001718 Homo sapi
43	556.2	10.9	340000	9	AF064245 Homo sapi
44	553.8	10.9	2020	9	HS1TSN1 AC144367 Papio anu
45	540.8	10.6	184526	2	

ALIGNMENTS

RESULT 1
AF132478
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999
DEFINITION Mus musculus Esei protein mRNA, complete cds.
ACCESSION AF132478
VERSION AF132478.1 GI:4378884
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Sengar, A.S., Wang, W., Bishay, V., Cohen, S., and Egan, S.B.
1 (bases 1 to 3723)
The EH and SH3 domain Esei proteins regulate endocytosis by linking
to dynamin and Eps15

JOURNAL	EMBO J. 18 (5), 1159-1171 (1999)
MEDLINE	99164083
PUBMED	10064583
REFERENCE	2 (bases 1 to 3723)
AUTHORS	Sengar,A.S., Wang,W., Cohen,S., Bishay,J. and Egan,S.E.
TITLE	Direct Submission
JOURNAL	Submitted (02-MAR-1999) Programs in Cancer & Blood
FEATURES	University/developmental Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, ON M5G 1X8, Canada Location/Qualifiers

```
source
1. .3723
  /organism="Mus musculus"
  /mol_type="rRNA"
  /db_xref="taxon:10090"
CDS
1. .3642
```

```

/codon_stat=1
/product="Ese1 protein"
/protein_id="AADI9746.1"
/db_xref="GI:437885"
/translation="MAORPFGSGLDWYATVVERBAKQDILSKETIAGTIPDPA
NFFPQSGLPQVLAQIWTALDAMNDGMDVSEYRANKLILKLGQIGQLSTLPPV
KQOPVAISSAPAFVIGIGIASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPAPVPLA
GAPYIQLPAPFAHPAAWTMPKSSSPSGPSQSLATKQKQSPVAPASAPGALVAT
QSSRIKYQLRNSHDKTSMGHLTGQOATILMQSLPAQALSIINLSDIDQDGLATV
BEFLIAMHLIDVAMSGQPLPYLPPEYIPSPRRIRSGSGMVISSSVDORLPPEP
SESDQOPEKTPYTPEDKKRPNFSGVSEILKRRALLQORKEORALQALPRAEPE
KEREOEOBAGQULEKOLEKORELEKOREBERKELERKEBAKRELKOROLEME
NRBOLLNORNKKEOEGTVLAKRRKTELFELELANDKHQOLEGKQODIRCLARTQOL
IESTKSNRELRIAEITLHQOOLESQOMQIGRIPEKQILSQOLKVQONSSLAHRSIL
LKRALKEELARQOLREOLDVEYERETRLQDEIVFNQOLKELRHISKQOLKORSI
BAARLKOKEORRKSLELKOKEDADRRQOEORQOEKQMLEHVQOEORPQRPEDEKL
EDSVRKAERBARPEMDQOSRLPHQOEPAKLAQAQPMTEKGLPTISAQSVYK
VYTRALPFSRSHDEITIQPQDILVWVDSQTEGPEVGLGELKGTGMFPAYATKIL
ENEPVPAKPVTLDTLSAPAKLALRETPAPLPVTSSESTPTNNMADSSVPMSSSN
PEEDNDMDTMAAOESLITVPSAQOLRORSAFTPATITQSSPSPVLQGGKGVGLQAQAL
YPMKAKDNHNLNFNSDVIITVLEQODMMFGVQOGKMPKSYVYKLISGVEKTS
DTGTPVSPASLKRVAVSPAKPAIIPGEETIANTYTSSEOGDITFOOGVIVITVKD
IMTGVGSGCVFSPNRYARILKDSBEGSNATGSLGKRLDEINQVIAVAAAGPEOLDIT
APGQILILRKKNPGSGMEGELQARCKKQOIGFNPANYKLLSPGTSKITPTPLPRTA
QPAVCQVIGMTDYTAQNQNDLAFSKQIILNVLNKEDPMMWKGESVGYVGLPSPNYVK
ITDMDPSQO"

```

BASE COUNT	1065 a	940 c	1026 g	692 t
ORIGIN				
Query Match	73.2%	Score 3721.4;	DB 10;	Length 3723;

Query Match	73.2%;	Score 3721.4;	DB 10;	Length 3723;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3722; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0

OY	259	ATGGCTCAGTTTCCACACCTTTGGGTGAGCTCGAATGTCGTGGCCATAACTGTGAG	318
Db	1	ATGGCTCAGTTTCCACACCTTTGGGTGAGCTCGAATGTCGTGGCCATAACTGTGAG	60
OY	319	GAAGGGCCACCATGACCAGACAGTTTCTTACCTGAAGCCGATAGCCGGATTATTA	378
Db	61	GAAGGGCCACCATGACCAGACAGTTTCTTACCTGAAGCCGATAGCCGGATTATTA	120
OY	379	GGTGATCAAGCGAGGAATTTTTTTCATCTGGGTAACTCAGCCCTCTTACACAA	438
Db	121	GGTGATCAAGCGAGGAATTTTTTTCATCTGGGTAACTCAGCCCTCTTACACAA	180
OY	439	ATATGGCGCTAGCGGACATGAATAAGATGGAAGATGATCAAGTGAATTTTCATA	498
Db	181	ATATGGCGCTAGCGGACATGAATAAGATGGAAGATGATCAAGTGAATTTTCATA	240
OY	499	GCCATGAAGCTTATCAACTGAAGCTACAAAGATACAGTCCCTCCCACTTCCCT	558
Db	241	GCCATGAAGCTTATCAACTGAAGCTACAAAGATACAGTCCCTCCCACTTCCCT	300
OY	559	GTCATGAACAACAGCAACAGTGGCTATTTTCCAGTGCACCGAATTTGGTATAGAGGAT	618
Db	301	GTCATGAACAACAGCAACAGTGGCTATTTTCCAGTGCACCGAATTTGGTATAGAGGAT	360
OY	619	GCTACGATCCACCACTCAGCTGTGCTCTGTGCCAATGAGGCTCAATTCAGTGT	678

Db	361	GCATGATGCCACCACTCACAAGCTGTGGCTCTGTGCGAATGGGCTCATTTCCAGTTGTT	420
Qy	679	GGAAATGTCCTCACCCCTTAGTATCTTCTGTCCCTTCAGACGAGTGCTCCCTGGCTAAC	738
Db	421	GGAAATGTCCTCACCCCTTAGTATCTTCTGTCCCTTCAGACGAGTGCTCCCTGGCTAAC	480
Qy	739	GGGGCTCTCCCGGCATACAGCCCTGCGCGTTGGCGGATCCTGTGAGCAATGGCCA	798
Db	481	GGGGCTCTCCCGGCATACAGCCCTGCGCGTTGGCGGATCCTGTGAGCAATGGCCA	540
Qy	799	AAGAGTCTTCTCTTCAGACAGATCTGTCTCAGGGGCACAAATTAAACATAAGTTACAGAG	858
Db	541	AAGAGTCTTCTCTTCAGACAGATCTGTCTCAGGGGCACAAATTAAACATAAGTTACAGAG	600
Qy	859	GCACAATCATTCGATGTGCGCCAGCGCCCTTCAGACAGCAATGGGCTGTGCTCACTCA	918
Db	601	GCAACAATCATTCGATGTGCGCCAGCGCCCTTCAGACAGCAATGGGCTGTGCTCACTCA	660
Qy	919	TCAAGGCTGAAATTCAGGCAAGTTATTTCAAGCAGCAACAAACTATAGATGGACAATTAA	978
Db	661	TCAAGGCTGAAATTCAGGCAAGTTATTTCAAGCAGCAACAAACTATAGATGGACAATTAA	720
Qy	979	ACAGGTCCTCCAGGCAAGAACTTATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT	1038
Db	721	ACAGGTCCTCCAGGCAAGAACTTATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT	780
Qy	1039	TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGCAGAGAAATTTATC	1098
Db	781	TCAATATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGCAGAGAAATTTATC	840
Qy	1099	CTAGCTATGCACTTAATTTGATGTTGGCCATGTCTGTGACGCCATGCGCCCGCTCTGCGCT	1158
Db	841	CTAGCTATGCACTTAATTTGATGTTGGCCATGTCTGTGACGCCATGCGCCCGCTCTGCGCT	900
Qy	1159	CCAGAAATACATCCCTCCCTCTTCCAGAAAGATTGGCTCCGGCAGTGGGATGTCCGTGATA	1218
Db	901	CCAGAAATACATCCCTCCCTCTTCCAGAAAGATTGGCTCCGGCAGTGGGATGTCCGTGATA	960
Qy	1219	AGCTCTTCTTCTGTGATTCAGAGGCTGCTGAGGAGCCGTGTGTCAGAGATGACAGCAG	1278
Db	961	AGCTCTTCTTCTGTGATTCAGAGGCTGCTGAGGAGCCGTGTGTCAGAGATGACAGCAG	1020
Qy	1279	CCAGAGAAAGAACTGCTCTGTGACATTTGGAAGATTAAGAGCGGAGAACTTTCAGCGAGGC	1338
Db	1021	CCAGAGAAAGAACTGCTCTGTGACATTTGGAAGATTAAGAGCGGAGAACTTTCAGCGAGGC	1080
Qy	1339	AGTGTGAGCTGTGAGAAAGCGCCGCAAGCGCTCTTGAAGCAGACGCGCAAGACAGAG	1398
Db	1081	AGTGTGAGCTGTGAGAAAGCGCCGCAAGCGCTCTTGAAGCAGACGCGCGCAAGACAGAG	1140
Qy	1399	CGGTTGGCTCAGCTGTGAGCGCGCCGACAGAGAGAGAAAGACGCGACAGAGCAG	1458
Db	1141	CGGTTGGCTCAGCTGTGAGCGCGCCGACAGAGAGAGAAAGAGCGCGACAGAGCAG	1200
Qy	1459	GAGGCGAAGCGCGAGCTGTGAGCTGTGAGAGAGACTTGAAGAACAGCGGAGCTGTGACGG	1518
Db	1201	GAGGCGAAGCGCGAGCTGTGAGCTGTGAGAGAGACTTGAAGAACAGCGGAGCTGTGACGG	1260
Qy	1519	CAGGGAAGAGAGAGAGAGAGAGAGATGAGAGAGGCGGAGCGCGCAAAACGGGAACTG	1578
Db	1261	CAGGGAAGAGAGAGAGAGAGAGAGATGAGAGAGGCGGAGCGCGCAAAACGGGAACTG	1320
Qy	1579	GAAAGGCGACACAACTTGAATGGGAAACGGAACCGGAGACAGAACTCTGAATCAGAGG	1638
Db	1321	GAAAGGCGACACAACTTGAATGGGAAACGGAACCGGAGACAGAACTCTGAATCAGAGG	1380
Qy	1639	AACAAAGAGCAGAGAGGACCGGTGTCTCTGAAGCAGAGAGAGAACTTGAATTTGAG	1698
Db	1381	AACAAAGAGCAGAGAGGACCGGTGTCTCTGAAGCAGAGAGAGAACTTGAATTTGAG	1440
Qy	1699	TTAGAAGCTTGAAATGACAAAGACATCAGTTAAGAGAAACTTCAGAGATATCAGGTGT	1758

Db	1441	TTAGAGCTCGAATGACAAAGATCAGCTAGAGAGAAATTCAGAGATATCAGGTG	1500
QY	1759	CGACTGCGAACCCAGAGGCAAGAAATTGAGACACGAACTAGAGACTAAGATT	1818
Db	1501	CGACTGGCAACCCAGAGGCAAGAAATTGAGACACGAACTAGAGACTAAGATT	1560
QY	1819	GCTGAATATCAACCACTTACAGACGAGTTGACGAAATTCACGAAATGCTTGAAGACTT	1878
Db	1561	GCTGAATATCAACCACTTACAGACGAGTTGACGAAATTCACGAAATGCTTGAAGACTT	1620
QY	1879	ATTCAGAGAAACAGATATCTCAGTACCGAGTTAAACAAAGTCACGACGAACTGATGAT	1938
Db	1621	ATTCAGAGAAACAGATATCTCAGTACCGAGTTAAACAAAGTCACGACGAACTGATGAT	1680
QY	1939	AGAGACTCGCTTCTTACCTCAAAAGACCTTGGAGCAAGAGAGCTGCCCGACAG	1998
Db	1681	AGAGACTCGCTTCTTACCTCAAAAGACCTTGGAGCAAGAGAGCTGCCCGACAG	1740
QY	1999	CTCCGGGAGCAGCTTGACGAGGTGAGAGAGACCAAGCTCAGAGATTGAT	2058
Db	1741	CTCCGGGAGCAGCTTGACGAGGTGAGAGAGACCAAGCTCAGAGATTGAT	1800
QY	2059	GTTTTCAACAAACCCAGCTGAGAGAACTGAGAGATACCTACCAACAGCACTCCAGAG	2118
Db	1801	GTTTTCAACAAACCCAGCTGAGAGAACTGAGAGATACCTACCAACAGCACTCCAGAG	1860
QY	2119	CAGAGGTCCTGAGAGGCAAGCCGACTGAGAGAAAGACGAGAGAGAGAGCTGGAG	2178
Db	1861	CAGAGGTCCTGAGAGGCAAGCCGACTGAGAGAAAGACGAGAGAGAGAGCTGGAG	1920
QY	2179	TTAGAGAAAGCAAAAGAAAGACGCTCAGAGACGAGTTGAGAAAGGAGCAAGCAATGGCTG	2238
Db	1921	TTAGAGAAAGCAAAAGAAAGACGCTCAGAGACGAGTTGAGAAAGGAGCAAGCAATGGCTG	1980
QY	2239	GAGCAATGACAGCAGAGAGAGACGACAGCCCGCCGAAACCCCAACAGAGAGACAGACTG	2298
Db	1981	GAGCAATGACAGCAGAGAGAGACGACAGCCCGCCGAAACCCCAACAGAGAGACAGACTG	2040
QY	2289	AAGAGGGAAGACAGTGTACAGAGAAAGAGCGGAGAGAGAGAGCCGGAATGCA	2358
Db	2041	AAGAGGGAAGACAGTGTACAGAGAAAGAGCGGAGAGAGAGAGCCGGAATGCA	2100
QY	2359	GACAAAGCAGAGTCGGCTTTTCCATCCGATCAGAGCCAGCTAACTGACCACCAAGCA	2418
Db	2101	GACAAAGCAGAGTCGGCTTTTCCATCCGATCAGAGCCAGCTAACTGACCACCAAGCA	2160
QY	2419	CCCTGCTTACCAACAGAGAAAGCCCGCTTACATTTTGTGACAGAGAGGTGTAAAGTG	2478
Db	2161	CCCTGCTTACCAACAGAGAAAGCCCGCTTACATTTTGTGACAGAGAGGTGTAAAGTG	2220
QY	2479	GTAATTTACCGAGCGCTGATCCCTTTAAATCCGAAGTCAAGATGAGATCAGCATCAG	2538
Db	2221	GTAATTTACCGAGCGCTGATCCCTTTAAATCCGAAGTCAAGATGAGATCAGCATCAG	2280
QY	2539	CCAGAGATATAGTATGATGATGATGAAACCAAGCTGAGAGCCAGAGATGCTTGAAGGA	2598
Db	2281	CCAGAGATATATGATGATGATGATGAAACCAAGCTGAGAGCCAGAGATGCTTGAAGGA	2340
QY	2599	GAGCTGAAAAGGAGACGGGATGTTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAAT	2658
Db	2341	GAGCTGAAAAGGAGACGGGATGTTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAAT	2400
QY	2659	GAGGTTCACATCCAGGCAAAACAGTGCAGATCGATCGCCCTGCCCCCAAACTG	2718
Db	2401	GAGGTTCACATCCAGGCAAAACAGTGCAGATCGATCGCCCTGCCCCCAAACTG	2460
QY	2719	GCTCTGCGTGAGACCCCTGCTCTTTGGCAGTGAATCTTCTGAGCCCTCCACAAACCC	2778
Db	2461	GCTCTGCGTGAGACCCCTGCTCTTTGGCAGTGAATCTTCTGAGCCCTCCACAAACCC	2520
QY	2779	AACAACTGGGAGACTTCAATTTCCAGTGGCCCAAGACTCAACGAGAGCCAGAAACG	2838
Db	2521	AACAACTGGGAGACTTCAATTTCCAGTGGCCCAAGACTCAACGAGAGCCAGAAACG	2580

QY	2839	GACAACTGGGATACAGTGGGCGGCTCAGCCTTCTGTACCCGTAACCTAAGTGTGGCAGATTA	2839
Db	2561	GACAACTGGGATACAGTGGGCGGCTCAGCCTTCTGTACCCGTAACCTAAGTGTGGCAGATTA	2640
QY	2839	CGGCAAGATCAGCCTTACCCCAAGCACAAGCACAAGTGGCTCTCCCAATCTCCGATCTG	2958
Db	2641	CGGCAAGATCAGCCTTACCCCAAGCACAAGCACAAGTGGCTCTCCCAATCTCCGATCTG	2700
QY	2959	GGCCAGGCTGAAAAGTGGAAAGGCTCAGCGCAAGCCTGTATCCCTGGAGACCAAA	3018
Db	2701	GGCCAGGCTGAAAAGTGGAAAGGCTCAGCGCAAGCCTGTATCCCTGGAGACCAAA	2760
QY	3019	AAAGCAACCACTTAATTTTAACAAAAGTAGCTCATCACGATCTGGAAACACAAAGAC	3078
Db	2761	AAAGCAACCACTTAATTTTAACAAAAGTAGCTCATCACGATCTGGAAACACAAAGAC	2820
QY	3079	ATGTGTGTGTTGGAGAAATTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAA	3138
Db	2821	ATGTGTGTGTTGGAGAAATTCAAGGTCAAGAGGTTGGTTCCCAAGTCTTACGTGAA	2880
QY	3139	CTCATTTCAAGGCCCCGTAAAGAAATCCACAAGCATCGATCTGGCCTTACTGAAGTCT	3198
Db	2881	CTCATTTCAAGGCCCCGTAAAGAAATCCACAAGCATCGATCTGGCCTTACTGAAGTCT	2940
QY	3189	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCCATTCCCGAGAGAGATT	3258
Db	2941	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCCATTCCCGAGAGAGATT	3000
QY	3259	ATTGCCATGTACACATACGAGAGTTCTGAGCAGAGAGATTTAACTTTCAGCAAGGGAT	3318
Db	3001	ATTGCCATGTACACATACGAGAGTTCTGAGCAGAGAGATTTAACTTTCAGCAAGGGAT	3060
QY	3319	GTTGATTTGTGTTACCAAGAAAGATGTGATCTGTGTGACGGGAACGGTGGGAGCAAGTCC	3378
Db	3061	GTTGATTTGTGTTACCAAGAAAGATGTGATCTGTGTGACGGGAACGGTGGGAGCAAGTCC	3120
QY	3379	GGAAGTCTTCCCTTCTAACTATGTAGAGCTTAAAGATTCAAGAGGCTCTGAACTGTGG	3438
Db	3121	GGAAGTCTTCCCTTCTAACTATGTAGAGCTTAAAGATTCAAGAGGCTCTGAACTGTGG	3180
QY	3439	AAAACAGGGAATTTAGAAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACAGCTCT	3498
Db	3181	AAAACAGGGAATTTAGAAAAAAAACCTGAATTTGCCAGGTTATTGCTTCTACAGCTCT	3240
QY	3489	ACTGGTCCCGAACAACCTACCTGGCTCTGTGGCAGCTGATTCTGATCCGGAATAAGAC	3558
Db	3241	ACTGGTCCCGAACAACCTACCTGGCTCTGTGGCAGCTGATTCTGATCCGGAATAAGAC	3300
QY	3559	CCAGGTGATGTGGGAAGAGAACTGCAAGCTTCAAGGGAAAAAGCGCCAGATNAGGTGG	3618
Db	3301	CCAGGTGATGTGGGAAGAGAACTGCAAGCTTCAAGGGAAAAAGCGCCAGATNAGGTGG	3360
QY	3619	TTTTCAGAAATTTATGTCAAACTTCTAAGCCCGGGAACAAGCAAAATCACCCCAACTGAG	3678
Db	3361	TTTTCAGAAATTTATGTCAAACTTCTAAGCCCGGGAACAAGCAAAATCACCCCAACTGAG	3420
QY	3679	CTACCCAAGACCGCAGTGCAGCCGACGAGTGTCCAGAGTGAATCGGAGATGAACGATTACCC	3738
Db	3421	CTACCCAAGACCGCAGTGCAGCCGACGAGTGTCCAGAGTGAATCGGAGATGAACGATTACCC	3480
QY	3739	GCCCAAGAACATGACGAACTTAGCTTTCAGCAAAAGCCAGATCATCAACGTCTCAACAG	3798
Db	3481	GCCCAAGAACATGACGAACTTAGCTTTCAGCAAAAGCCAGATCATCAACGTCTCAACAG	3540
QY	3799	GAGGACCCGGAAGTGGAGAAAGAGAAAGTCAAGTGGGCAAGTGGGCTCTTCCCAATCCAAAT	3858
Db	3541	GAGGACCCGGAAGTGGAGAAAGAGAAAGTCAAGTGGGCAAGTGGGCTCTTCCCAATCCAAAT	3600
QY	3859	TATGTAAAGCTGACCAACAGATGAGCCCAAGCAGCAATGAATCATATGTTGTCAATCC	3918
Db	3601	TATGTAAAGCTGACCAACAGATGAGCCCAAGCAGCAATGAATCATATGTTGTCAATCC	3660

[illegible]

BASE COUNT	1448 a	1298 c	1415 g	984 t	
Query Match	71.6%	Score 3639.4	DB 10	Length 5145	
Best Local Similarity	100.0%	Pred. No. 0			
Matches 3640	Conservative 0	Mismatches 1	Indels 0	Gaps 0	
Qy	259	ATGGCTCAGTTTCCCAACACCTTTTGGGAGTACCTGGATGTCCTGGGCCATAACTGTGGAG	318		
Db	1	ATGGCTCAGTTTCCCAACACCTTTTGGGAGTACCTGGATGTCCTGGGCCATAACTGTGGAG	60		
Qy	319	GAAGGGCCAGCATGATACACAGTCTTCTAGCTGAAGCCGATGCGGAGTTTATTACT	378		
Db	61	GAAGGGCCAGCATGATACACAGTCTTCTAGCTGAAGCCGATGCGGAGTTTATTACT	120		
Qy	379	GGTATCAAGCGAGAACTTTTTTCCAACTCTGGATTACCTAGCCCTGTCTTAGCAAA	438		
Db	121	GGTATCAAGCGAGAACTTTTTTCCAACTCTGGATTACCTAGCCCTGTCTTAGCAAA	180		
Qy	439	ATATGGGCGCTAGGGGACATGAATAACATGGAAGATGATCAAGTGAATTTTCCATA	498		
Db	181	ATATGGGCGCTAGGGGACATGAATAACATGGAAGATGATCAAGTGAATTTTCCATA	240		
Qy	499	GCCATGAAGCTTATCAAACTGAGCTACAAGATATACGCTCCCTCCACCTTCCCTCT	558		
Db	241	GCCATGAAGCTTATCAAACTGAGCTACAAGATATACGCTCCCTCCACCTTCCCTCT	300		
Qy	559	GTCAATGAACAGCAACAGTGGCTATTCCAGTGACACCAAGATTGTATAGAGGATT	618		
Db	301	GTCAATGAACAGCAACAGTGGCTATTCCAGTGACACCAAGATTGTATAGAGGATT	360		
Qy	619	GCTAGCATGCCACAATCAACAGCTGTGCTCTGTGCCAATGAGGCTCATTCAGTTGT	678		
Db	361	GCTAGCATGCCACAATCAACAGCTGTGCTCTGTGCCAATGAGGCTCATTCAGTTGT	420		
Qy	679	GGATATGCTCCACCCCTAGTATCTTCTGCTCCAGAGAGAGTGCCTCCCTGGCTAAC	738		
Db	421	GGATATGCTCCACCCCTAGTATCTTCTGCTCCAGAGAGAGTGCCTCCCTGGCTAAC	480		
Qy	739	GGGGCTCCCTCCCTCATACACAGCTCTGCTGCTGGTGCATCTCTGACCCACATGGCCA	798		
Db	481	GGGGCTCCCTCCCTCATACACAGCTCTGCTGCTGGTGCATCTCTGACCCACATGGCCA	540		
Qy	799	AAGATTTCTTCTTACAGATCTGGTTCAGGGTCACAATTTAAACATAAGTTACAGAA	858		
Db	541	AAGATTTCTTCTTACAGATCTGGTTCAGGGTCACAATTTAAACATAAGTTACAGAA	600		
Qy	859	GCACAATCATTCGATATGTCGACAGGGCCCTCCAGACAGAAATGGGCTGTGCTCAGTCA	918		
Db	601	GCACAATCATTCGATATGTCGACAGGGCCCTCCAGACAGAAATGGGCTGTGCTCAGTCA	660		
Qy	919	TCAAGGCTGAATATCAGGCAAGTATTCAACAGCCACGACAAACATATAGTGAACATTA	978		
Db	661	TCAAGGCTGAATATCAGGCAAGTATTCAACAGCCACGACAAACATATAGTGAACATTA	720		
Qy	979	ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCTCAGCTGACT	1038		
Db	721	ACAGGTCCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCTCAGCTGACT	780		
Qy	1039	TCAATATGGAATCTTTCTGACATTTGATCAAGTGAATACTCATGCAAGAAATTTATC	1098		
Db	781	TCAATATGGAATCTTTCTGACATTTGATCAAGTGAATACTCATGCAAGAAATTTATC	840		
Qy	1099	CTAGCTATGCACTAAATGATGTTGCCATGCTGATGCACTGCGCCGCTCTGCTGCT	1158		
Db	841	CTAGCTATGCACTAAATGATGTTGCCATGCTGATGCACTGCGCCGCTCTGCTGCT	900		
Qy	1159	CCAGAATACATCCCTCTCTTCCATCAGAAAGATTGCTCCGCGAGTGGAGTCCGTCTATA	1218		
Db	901	CCAGAATACATCCCTCTCTTCCATCAGAAAGATTGCTCCGCGAGTGGAGTCCGTCTATA	960		
Qy	1219	AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGTGTGACAGAGATGACAGAG	1278		

Qy 3439 AAACAGGAGGATTGAGAAAAAACTGAAATTCGCCAGTTATGCTTCTAGCGTCT 3498
 Db 3181 AAAACAGGAGGATTGAGAAAAAACTGAAATTCGCCAGTTATGCTTCTAGCGTCT 3240
 Qy 3499 ACTGGTCCCGAACAACCTACCTGGCTCTCTGGGCACTGATTCGATCCGAAAAAGAAC 3558
 Db 3241 ACTGGTCCCGAACAACCTACCTGGCTCTCTGGGCACTGATTCGATCCGAAAAAGAAC 3300
 Qy 3559 CCAGGTGATGGTGGAGAGAGAACTGCAAGCTGAGGAAAAAAGCGGACATGAGGAG 3618
 Db 3301 CCAGGTGATGGTGGAGAGAGAACTGCAAGCTGAGGAAAAAAGCGGACATGAGGAG 3360
 Qy 3619 TTTCCAGCAAAATTAATGCTAACTTCTAAGCCCGGAGAACAAATATCACCCTGAG 3678
 Db 3361 TTTCCAGCAAAATTAATGCTAACTTCTAAGCCCGGAGAACAAATATCACCCTGAG 3420
 Qy 3679 CTACCAAGACCGCAGTGCAGCCAGCAGTGTGCCAGTGTGATCGGATGATTAACACC 3738
 Db 3421 CTACCAAGACCGCAGTGCAGCCAGCAGTGTGCCAGTGTGATCGGATGATTAACACC 3480
 Qy 3739 GCCCAAGACGATGAGAACTAGCTTCCAGCAAAAGCGGACATGATCAAGCTCTCAACAG 3798
 Db 3481 GCCCAAGACGATGAGAACTAGCTTCCAGCAAAAGCGGACATGATCAAGCTCTCAACAG 3540
 Qy 3799 GAGGACCCGAGCTGTGGAGAGAGAGAGTCAAGTGGGCAAGTTGGCTCTTCCATCCAT 3858
 Db 3541 GAGGACCCGAGCTGTGGAGAGAGAGAGTCAAGTGGGCAAGTTGGCTCTTCCATCCAT 3600
 Qy 3859 TATGTAAAGCTGACACACAGATGAGACCCGACCCAGCAATG 3899
 Db 3601 TATGTAAAGCTGACACACAGATGAGACCCGACCCAGCAATG 3641

RESULT 3
 AF114488 5381 bp mRNA linear PRI 16-JUL-2002

LOCUS Homo sapiens intersecin short isoform (ITSN) mRNA, complete cds.
 DEFINITION AF114488
 ACCESSION AF114488.1 GI:4808824
 VERSION
 KEYWORDS
 ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 5381)
 AUTHORS Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

ABSTRACT Pucharcos, C., Fuentes, J.J., Casas, C., de la Luna, S., Alcantara, S.,
 Arbones, M.L., Soriano, E., Estivill, X. and Pritchard, M.
 Multivalent binding of human intersecin (ITSN), a putative
 differentiating neurons and overexpressed in Down syndrome

JOURNAL Eur. J. Hum. Genet. 7 (6), 704-712 (1999)
 MEDLINE 99415290
 PUBMED 10482960

REFERENCE 2 (bases 1 to 5381)
 AUTHORS Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.
 TITLE Direct Submission

JOURNAL Submitted (16-DEC-1998) Medical and Molecular Genetics Center,
 Cancer Research Institute, V/Hospital de DIO, Avia,
 Castelldefels km. 2,7, Barcelona 08907, Spain

FEATURES
 source location/Qualifiers
 1..5381

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1-q22.2"

1..5381
 /gene="ITSN"
 /gene="ITSN"
 /codon_start=1
 /product="intersecin short isoform"
 /protein_id="AAD29953.1"

gene
 CDS

/db_xref="GI:4808825"
 /translation="MAQPPFGSGSLDIAITVEERAKHQOFSLKPSIGTIGDQA
 RNFPSGLPQVLAQIMALADNMNDNRMOVERSIAMKLIKLGQVLPALPMA
 KOQVPAISSAPAFMGISAMPPLTANAPPMSPVPGMSPTVSVAPPAAPLAN
 GAPVIOPLPAFAHPATLFKSSFSRSGSGSLTLOQAQSFDAVSPVAVMAVP
 OSSRLKYROLFNSHDKTMSGLTGPOKRLIMQSLTQALSWLSDIDQKTLA
 EEPFLAMHLIDVAMSGPLPPLVPEPIPSFPRVRSGLSVISSTVQVRLPEEY
 LEDEOLLEKKLPVTFEDKRENERENILLEKROQLLEQKREORLQLEBAEY
 RKREEROERKQLEIKOLEKORLEEREREERREIRREARKEERORLEME
 RNRROELNQRNEQEDIVYIAKKKLELELELELELELELELELELELELE
 EISTNSRELRILAEITLHLOQLOESQMLGRLLPEQIILNDLQKQVONSLEHDSLV
 TLKRALFAKELARQHLDDLEVEKERSLQLEIDFNNQLKELEIHNQLOQKQS
 MEAEERLKOKEORLELEKORKEBAQRAQERKQMLVQOEDEHORPKLREBY
 KRRESVKQKQEEKQEOADKGLRPHQOEBAQRAQERKQMLVQOEDEHORPKLREBY
 KRVYVRLVPEERSESHDEITIOGDIYMWGEMWDESGTEGPGMLGSELKGTGMPFA
 NYAKLIPENEPVAPVRYVTSTSPAPAKLALRETPALATSSPSPTPNMADPSST
 WPTSTNEKPEPTDMDANAAPSLVPEAGLOQASATPATAVTSSPVPGQEKYE
 GLQALVLPWRARQNDNLNPKNDVITVLQODMWFQVQGVKPKSVKLIISGP
 IRKSTSDSGSSPSPLKRVASPAAPVPSGEBFLAMYSESQGLDLPFOQDVL
 VTKKDDPMWTVGDKAGVPSNVYRLKDESGSTAGTSLGKKEPIAVIATYAT
 GPEQLTLAPOLILIRKKNPGMMEGLQARKKROIQWFPANVYKLISPGTSKITPT
 EPRKSTLAALVCVIGNYDTTAQNDDELAENKQIIVLNKEDPDMKKEVNGVGLF
 PSNVYKLTMDPSQ"

BASE COUNT 1629 a 1192 c 1304 g 1256 t
 ORIGIN
 Query Match 69.4%; Score 3529.4; DB 9; Length 5381;
 Best Local Similarity 83.7%; Pred. No. 0;
 Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;
 Qy 7 GAGGAGAGTGAAGAGCGGCGGAGGAGGCGGAGCTTGTGCTCCGTAGTACCGCGCT 66
 Db 27 GAGGAGAGTGAAGAGCGGCGGAGGAGGAGGAGCTTGTGCTCCGTAGTACCGCGCT 86
 Qy 67 CGCAAGGAGATCCCGAGCGGCTCCGAGACGCGCGGAGGAGGAGCGGCGGCGG 126
 Db 87 CGCAGGAGAGATCCCGAGCGGCTCCGAGACGCGCGGAGGAGGAGCGGCGG 135
 Qy 127 GGGATGATGCGGCGGCTGCGGAGCTCGGCGCTTCTGCGGCGGCGGCTGCGGCTG 185
 Db 136 GGGATGATGCGGCGGCTGCGGAGCTCGGCGCTTCTGCGGCGGCGGCTGCGGCTG 195
 Qy 186 TTTGTGAGAGGAGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 245
 Db 196 TTTGTGAGAGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 255
 Qy 246 AGCTAATGAAACATGCTGCTCCCAACCTTTGGTGTGAGCTGCTGCTGCTGCTG 305
 Db 256 AAGTAAAGAAACATGCTGCTCCCAACCTTTGGTGTGAGCTGCTGCTGCTGCTG 315
 Qy 306 CATACCTGTGAGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 365
 Db 316 CATACCTGTGAGAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 375
 Qy 366 GGGATTTATTAATCTGTGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 425
 Db 376 TGAATTCATTAATCTGTGATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 435
 Qy 426 TGTCTTACACAAATATGAGCGCTAGCGGAGCATGATGAGGAGGAGGAGGAGGAG 485
 Db 436 TGTCTTACACAGATATGAGCGCTAGCGGAGCATGATGAGGAGGAGGAGGAGGAG 495
 Qy 486 GGAATTTTCCATACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 545
 Db 496 GGAATTTTCCATACCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 555
 Qy 546 CACACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 605
 Db 556 TGCACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 615
 Qy 606 TATAGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 665
 Db 616 TATAGAGGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 675

Db 2836 GTGGCCCAACGACGATGAGAAACGAAACGATTAATGAGATGATGGACGCCCA 2895
Qy 2865 GCCTTCTGACCGTACTAGTGTGCGCAGTTACGCGAGAGATCAGCTTTTACCCAGC 2924
Db 2886 GCCCTCTCAACCGTCCAAAGTCCGCGCAGTTAAAGCAGAGTCCGCTTTACTCCAGC 2955
Qy 2925 CACAGCCATGCGTCTCCCATCTCCGCTCGTGGCGCAGAGGTGAAAAGTGGAGGCT 2984
Db 2966 CACGCGCATGCGTCTCCCGTCTCTGCTAGGCGAGGTGAAAAGTGGAGGCT 3015
Qy 2985 ACAAGCGAAGCCCTGTATCCCTGGAGAGCCAAAAAGACAACCACTTTAAATTTTAA 3044
Db 3016 ACAAGCTAAGCCCTATATCTTGGAGAGCCAAAAAGACAACCACTTTAAATTTTAA 3075
Qy 3045 AAGTACGTATACCGTCTTGAGAACAGAACAGATGTGTGTGTTTGGAGAGTTCAAG 3104
Db 3076 AAATGATGTATCAACCGTCTTGAGAACAGAACAGATGTGTGTGTTTGGAGAGTT 3135
Qy 3105 TCAGAAAGGTGTGTTCCCAAGTCTTAGTGAATCACTTTCAAGGCGCGTAAAGAAATC 3164
Db 3136 TCAGAAAGGTGTGTTCCCAAGTCTTAGTGAATCACTTTCAAGGCGCGTAAAGAAATC 3195
Qy 3165 CACAAGCATCATCTGACCCCTACTGAAAGTCTGTAGTCTTAAAGAGAGTGGCTTCCC 3224
Db 3196 TACAAGCATCATCTGATTTCTTGAGAGATCTGTAGTCTTAAAGCAGATAGCTCTCC 3255
Qy 3225 GAGCGCCAAAGCCGATTTCCCGAGAAAGTTTATGCCATGACATACATACGAGATTC 3284
Db 3256 AGCAGCCAAAGCCGATTTCCCGAGAAAGTTTATGCCATGATCACTTACGAGATTC 3315
Qy 3285 TGAGAAAGAGATTTAAGCTTACAGAGGGAGTGTGTTGTGTACCAAGAAAGATG 3344
Db 3316 TGAGAAAGAGATTTAAGCTTACAGAGGGAGTGTGTTGTGTGTACCAAGAAAGATG 3375
Qy 3345 TGACTGTGAGACGGGAAACGTTGAGACAGAACTCCGAGTCTTCCCTTAACTATGTAG 3404
Db 3376 TGACTGTGAGACGGGAAACGTTGAGACAGAACTCCGAGTCTTCCCTTAACTATGTAG 3435
Qy 3405 GCTTAAAGATTCAGAGGCTCTGGAACGCTGTGGGAAAAAGAGGATTTAGAAAAAAC 3464
Db 3436 GCTTAAAGATTCAGAGGCTCTGGAACGCTGTGGGAAAAAGAGGATTTAGAAAAAAC 3495
Qy 3465 TGAATTTCCAGGTTATGCTTCTAGCTGTCTGCTCCGGAACAATCAACCTGAGC 3524
Db 3496 TGAATTTCCAGGTTATGCTTCTAGCTGTCTGCTCCGGAACAATCAACCTGAGC 3555
Qy 3525 TCTTGGGAGCTGATTTCTGATCCGGAAGAAAGAACCGAGTGTGTGGGAGAGAACT 3584
Db 3556 CCTGTGAGCTGATTTTGTATCCGAAAAAGAAACCAAGTGTGTGGGAGAGAACT 3615
Qy 3585 GCAAGCTGAGGAGAAAAAGCCGAGATAGGCTGTTCCTTCCAGCAATATATGCAACTTCT 3644
Db 3616 GCAAGCAAGTGGGAAAAAGCCGAGATAGGCTGTTCCTTCCAGCAATATATGCAACTTCT 3675
Qy 3645 AAGCCCGGAAACAACCAAAATCAACCCCACTGAGTACCAAGACCGAGTGCACACG 3704
Db 3676 AAGCCCTGGAGACGACCAAAATCACTCAACAGACGACCTTAAGTAAACGATTAAGCG 3735
Qy 3705 AGTGTGCAAGTGTATCGGATGTACATTAACACCCGCAAGACGATGACGATAGCTT 3764
Db 3736 AGTGTGCAAGTGTATCGGATGTACATTAACACCCGCAAGACGATGACGATAGCTT 3795
Qy 3765 CAGCAAGGCGCATCATCAACGCTCTCAACAAAGAGAGACCCGGAATGTGTGAAAGAGA 3824
Db 3796 CAACAAGGCGCATCATCAACGCTCTCAACAAAGAGAGACCCGGAATGTGTGAAAGAGA 3855
Qy 3825 AGTCAAGTGGGCAAGTGGGCTCTTCCCATCAATATATGTAAGCTGACCAAGACATGGA 3884
Db 3856 AGTCAAGTGGGCAAGTGGGCTCTTCCCATCAATATATGTAAGCTGACCAAGACATGGA 3915
Qy 3885 CCCCAGCAGCAATGATCATATGTGTGCATCCCCCTCAGGCTTGAAGTCTCTCAA 3944

Db 3916 CCAAGCGAGATGATCATATGTGTGCATCCCCCTCAGGCTTGAAGTCTCTCAA 3975
Qy 3945 GAGACCACTATCCCATATCACTGCCCAGAGAGATGATGGAGATGAGACCTTATCATG 4004
Db 3976 GAGACCACTATCCCATATCACTGCCCAGAGAGATGATGGAGATGAGACCTTATCATG 4035
Qy 4005 TGACTTGCAGATATCACTATCTGCTTGTGATGAGAGAACTCACTGAGACAGTGT 4064
Db 4036 TGACTTGCAGATATCACTATCTGCTTGTGATGAGAGAACTCACTGAGACAGTGT 4095
Qy 4065 ACCTCATTTGACCTTATGATGATGATGAGATGAGATGATGATGATGATGATGATG 4124
Db 4096 ACCTCATTTTAACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 4155
Qy 4125 AAGC--AAATGGAACGACAGGAGTGGGCTCTTGTGGGCTTCTTCTAGTCTC 4182
Db 4156 GAGCAAAATTTACAAAAACACAGGAGTGGGCTCTTGTGGGCTTCTTCTAGTCTC 4215
Qy 4183 AGACTGACCGG--CCCGGCTTCAACAGGCGCTTCAATAGTTTAAAGATTTTAA 4241
Db 4216 AAATGACTTCCCGACCTTGTGACAGAGTCTTCAATAGTTTAAATTTTAA 4275
Qy 4242 TGTGTATTTTAACTTTTAAATTTAAATCTCAATCAATTAATTTCTTGTGCTT 4301
Db 4276 TATATATTTTAACTTTTAAATTTAAATTTAAATTTAAATTTTGTGCTT 4335
Qy 4302 TACAAAAACCCCATATCAAGAGTGTCTGTGCGAGATTTAAATGCTGTCCGG 4361
Db 4336 TGCAAAAAAGCCCATATCAAGAGATG---CTGCATGTGTATTTAAATTTGTCCAA 4390
Qy 4362 CGTACCGTAAATGAGAGCTTGTGTACCTTGTGCGTGTGCTGCTTCCCAACAT 4421
Db 4391 ATGTCCATTAATCTGAGACTGTATGATTTTATTTTGTGCTGCTTCAACATTAAT 4450
Qy 4422 TGTGTAGTTGGGCTGTCTCC---TGCCTAGACACAGAGAGATGGTGTACTGT 4477
Db 4451 TGTGCAGTTTGGGCTTGTCTTCCCTTACATAGAGAGTGCAGAGATGTACTGTCTGT 4510
Qy 4478 TTTGAAATGTGTATGTATGATGAGCTGACTATGGAAGGGCTTATCTGTCTGTACC 4537
Db 4511 TTTTAAAGCCTATA--GATGAGCCCAATTTAAAGGAGGTGTGTGTGTGTGTGT 4569
Qy 4538 ATCACTGTATCC--TGTGCGCATGTACATCTGTACGAGAGAAATAGCT--CTTCTCAT 4595
Db 4570 ATCACTGTATCTTGTGATGATGTATATCACTCTGTATCAATTAATTAATTTCTTCA 4629
Qy 4596 GGTAAACCCACACCGGTATCAAGTGTCTCATCTACT--GCATTCATTTTACTTGCAC 4653
Db 4630 TGGCAAGCTATTAATCTTGTATGATGCTGTATATTAATTTTAAATTTTGTGCAC 4689
Qy 4654 AGTACCTTGTATGACCTGAGAGAGAC--CCATGTTCCGTTGTGTCTGATGTACT 4712
Db 4690 AGTACCTTGTATGACCTGAGAGAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGT 4749
Qy 4713 AGTGTGCGGCTGT 4772
Db 4750 GGTGTAGT 4809
Qy 4773 TAAAGCGCAACTGCCAGGCGGTATGATGATGATGATGATGATGATGATGATGATG 4827
Db 4810 GTAGCAAC--ACCATGAGGTGTGTATGATGATGATGATGATGATGATGATGATGATG 4868
Qy 4828 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4881
Db 4869 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4928
Qy 4882 CTTCGTGCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4941
Db 4929 TTTCTTGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4987
Qy 4942 TTTTAACTAGACTGTGGGGTGTCTACAGATTAATTAATGAAATGAGCTCTGTGCTG 5001
Db 4988 TTTTAACTAGACTGT--GAAATTTCTACAGATTAATTAATGAAATGAGCTCTGTGCTG 5046

QY 5002 TGTGTTACTGTCCTAGCTGAAGCCGTGTGCT---CCTTAAATTATGTTGAGAGT 5058
|||||
DB 5047 TGTGTTT-AGATATCTCTAGCTGAAGCCCTGTTGTTCTTTTAAACACTGATTGGAAGC 5105
|||||
QY 5059 CCGGAGAGAGATT 5071
|||||
DB 5106 TCTCATATAAAAT 5118
|||||

RESULT 4
AF127798 4025 bp mRNA linear ROD 28-JUN-1999
LOCUS Rattus norvegicus EH- and SH3-domain containing protein EHS1 mRNA,
DEFINITION complete cds.
ACCESSION AF127798
VERSION AF127798.1 GI:4835852
KEYWORDS
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 4025)
AUTHORS Okamoto,M., Schoch,S. and Sudhof,T.C.
TITLE EHS1/intersectin, a protein that contains EH and SH3 domains and
binds to dynamin and SNAP-25. A protein connection between
exocytosis and endocytosis?
JOURNAL J. Biol. Chem. 274 (26), 18446-18454 (1999)
MEDLINE 99303609
PUBMED 10373452
REFERENCE 2 (bases 1 to 4025)
AUTHORS Okamoto,M., Schoch,S. and Sudhof,T.C.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-1999) Center for Basic Neuroscience and HMMI,
Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX
75235, USA

FEATURES
source location/Qualifiers
1. 4025
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
41..3694
/codon_start=1
/product="EH- and SH3-domain containing protein EHS1"
/protein_id="AAD30271.1"
/db_xref="GI:4835853"

CDS
1
translating="MAQPTPGGSLDIWATVERAKHDQFOSLKPISGFTIGDOA
RNFPSGSLPQPLAIVMALDMNKDGRMDQVEFSIAMLILKLOGYOLPALPVM
KOOPAISSAPAPFGAGMGPPLTAVAPVWGSIPVWGSPSLVSVPVAPVPLAN
GAPVIOPLAPAPAPATLPKSSSFSRSGSLTKQAKOSFVYASAPAAEYAVP
OSSRLTKROLFNSHDKTMSGHILGFQARTILQSSLPQALASINLSDIDDKLTA
BEFILAMHLDIVMSGQPLPVLPEYIIPSPFRVRSQSGMVISSSADQLEPES
SEDOVEKRLPTEFEDKRENFERNGLLEKROALEQCKEDELRLQALRAQER
KEREROERKRLLEKOLEKORLELREERREKREIERRAARELROLOMER
NRROLLTQBNKDOGIIVLYKARKRLLEALDKHOLEGKLODRLCRLATGROE
LESTNKSRELRATITHTLOOQLOESQMGRIIPERKQLSDLKVOYQNSLHDSILT
LKGALFAKELAQQLREQLDEVEKSTRSLQELIVFNQQLKELRIHSKQLOQKRSI
EAKRLKQERKSLLEKQKEGQRVQERDKQOEHVQKQMGVYKLIISGVK
EDSVKKEAEERAKPEVDKQSLFPHQEPAPKPAQAPPTTEKGLTISAGESAKV
YRALYFESRSHDEITIOGDIVWKGWVESQTEGEGKPGKGTGFPANVA
EKIPENETPTAKPVTLTSLAPALALRETPALPLVTSSESTTPNNKADSSWPS
SNEKEPTDMNTWAAQPSLTITPSAGQLKRSATTPPATRTGSSPSPVQGEKVGLO
AOLYPMRAKQDNLNFKNSDITVLEODMMWFGEVQKQMGVYKLIISGVK
KSDWITGTEAPSSLSKRVASPAKAPALPEEFVAMYTESSSHGDLTPOGHVIVTK
KDGMWGTGTEGTSVFPNSNYRLKDSBSGTAGTGLSEKDEIAOVASTATGPE
OULAPGOLILIRKKNPGSMWEGELQARKROIGFPMANTYKLSLPGSKITPELP
KVAOVAVCOVIGMTDYTAQNDELAFSGQIINVLSKEDPDMWKGVSQVGLRPSN
YKLTDMPSQO"

BASE COUNT 1147 a 1009 c 1123 g 746 t
ORIGIN
Query Match

69.2%; Score 3517.2; DB 10; Length 4025;

Best Local Similarity 93.4%; Pred. No. 0;
Matches 3734; Conservative 0; Mismatches 243; Indels 21; Gaps 5;
QY 219 GAGATGAGCGCTGCATCAGCAAGTGAACGTATATGAACATGCTCACTTCCACACC 278
|||||
DB 1 GAGATGAGCGCTGCATCAGCAAGTGAACGTATATGAACATGCTCACTTCCACACC 60
|||||
QY 279 TTTGGTGTAGCCGTGAGTGTCTGGGCCATACTGTGAGGAAAGGCGCAACATGACCA 338
|||||
DB 61 TTTGGTGTAGCCGTGAGTGTCTGGGCCATACTGTGAGGAAAGGCGCAACATGACCA 120
|||||
QY 339 GCAGTTCCTTACCTGAGCCGATAGCCGAGATTTATCTGTGATCAAGCAAGAACTT 398
|||||
DB 121 GCAGTTCAGAGATTTGAAGCCGATATCTGATATCTGATCTGATCAAGCAAGAACTT 180
|||||
QY 399 TTTTTCATCTGGGTTACTCTAGCCTGTCTTACCAAAATATGCGCTAGCGACAT 458
|||||
DB 181 CTTTTCATCTGGGTTACTCTAGCCTGTCTTACCAAAATATGCGCGCTAGCGACAT 240
|||||
QY 459 GAATACGATGGAAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTATCAACT 518
|||||
DB 241 GAATACGACGGAAGATGATCAAGTGAATTTTCCATAGCCATGAAGCTTATCAACT 300
|||||
QY 519 GAAGCTACAGGATATACAGTCCCTCCACACTTCCCTGTATGAAACAGCAACACT 578
|||||
DB 301 GAAGCTGACAGGATATACAGTCCCTCCGCACTTCCCTGTATGAAAGCAGCAGCAGC 360
|||||
QY 579 GCGTATTTTCAGTCAACGACATTTGTATGAGAGGATTTGCTAGCATGCCACTCAC 638
|||||
DB 361 GCGCATCTCTAGTACACGACGTTGTATGAGAGGATTTGCTAGCATGCCACTCAC 420
|||||
QY 639 AGCTGTGCTCTGTGCAATGGGCTCATTCATCTGTTGAAGTCTCCACCTTACT 698
|||||
DB 421 AGCTGTGCTCTGTGCAATGGGCTCATTCATCTGTTGAAGTCTCCGCTTACT 480
|||||
QY 639 ATCTTCTGCTCTCAGACAGCAGTCCCTCCCTGCTAAACGAGGCTCCCTCCGTATACA 758
|||||
DB 481 ATCTTCTGCTCTCAGACAGCAGTCCCTCCCTGCTAAACGAGGCTCCCTCTGTATACA 540
|||||
QY 759 GCGTGTGCTCTGTGCGCATCTGACGCCACATGGCCMAAGATCTTCTCTTACGAG 818
|||||
DB 541 GCGTGTGCTCTGTGCGCATCTGACGCCACATGGCCMAAGATCTTCTCTTACGAG 600
|||||
QY 819 ATCTGTCCAGGATCACAATTAAACATTAAGTTACAGAAAGCAATCATTTGATGTCG 878
|||||
DB 601 ATCTGTCCAGGATCACAATTAAACATTAAGTTACAGAAAGCAATCATTTGATGTCG 660
|||||
QY 879 CAGCCGCCCTCAGACAGCAAGATGGGCTGTGCTCAGTCATCAAGGCTGAATAACAGCA 938
|||||
DB 661 CAGCCGCCCTCAGACAGCAAGATGGGCTGTGCTCAGTCATCAAGGCTGAATAACAGCA 720
|||||
QY 939 GTTATTAACAGCAGCAAACTATGATGACATTAACAGGTCCCGACGACAGAAC 998
|||||
DB 721 GTTATTAACAGCAGCAAACTATGATGACATTAACAGGTCCCGACGACAGAAC 780
|||||
QY 999 TATTTCTATGCAATCAAGTTTAAACCCAGGCTCAGTGGCTTCAATATGAAATCTTTTGA 1058
|||||
DB 781 TATTTCTATGCAATCAAGTTTAAACCCAGGCTCAGTGGCTTCAATATGAAATCTTTTGA 840
|||||
QY 1059 CATTTATCAAGATGAAAGTCACTGACGAAAGAAATTTATCTAGCTATGACCTTAATGA 1118
|||||
DB 841 CATTTATCAAGATGAAAGTCACTGACGAAAGAAATTTATCTAGCTATGACCTTAATGA 900
|||||
QY 1119 TGTTCATCTGTGTCAGCACTGCGCCGCTCTGCTCCAGAAATCATCTCTCTTC 1178
|||||
DB 901 TGTTCATCTGTGTCAGCACTGCGCCGCTCTGCTCCAGAAATCATCTCTCTTC 960
|||||
QY 1179 CTTCAAGAGTTGCTCTCCGACAGTGGATGTCCTCATTAAGCTTTCTTGTGAGATCA 1238
|||||
DB 961 CTTTAAAGAGTTGCTCTCCGACAGTGGATGTCCTCATTAAGCTTTCTTGTGAGATCA 1020
|||||
QY 1239 GAGGCTGCTGAGAGCGCTGCTGACAGATGACAGCAGCAGAGAAACTGCTGT 1298
|||||

Db	1021	GGGGCTCCCCGAGGAGCCATCTCTCAGAGAGATGACGACGAGGTGGAAAAAGAAAGCTGCTCT	1080
Qy	1239	GACATTTGAAAGATAAAGACCGGAGAACTTCGAGCGAGGCAAGTGTGAGCTTGAGAAACG	1358
Db	1081	GACATTTGAAAGATAAAGACCGGAGAACTTCGAGCGAGGCAACCTTGAGCTTGAGAAACG	1140
Qy	1339	CGCCAAAGCGCTCTTGAGAGCAGCAGCCGCAAAAGACAGAGCGGTTGGCTCAGCTTGAGCG	1418
Db	1141	CAGGAGAGCGCTCTTGAGAGCAGCAGCCGCAAAAGACAGAGCGGCTGAGCTCAGCTTGAGCG	1200
Qy	1419	CGCCAGAGGAGAGAAAGACCGGAGCGCCACAGAGAGAGGAGCGCAAGCGGACCTGGA	1478
Db	1201	CGCCAGAGAGAGAAAGACCGGAGCGCCACAGAGAGAGGAGCGCAAGAGCAGCTTGGA	1260
Qy	1479	GCTGAGAAAGCAGCTTGAGAAAGCAGCGGAGCTTGAGCGGACGCGAGAGAGAGAGAG	1538
Db	1261	GCTGAGAAAGCAGCTTGAAAGAAAGCAGCGGAGCTTGAGAGGACAGAGAGAGAGAGAG	1320
Qy	1539	GAAAGAGATCCGAGAGCGCGAGGCGCGCAAAACGGGAACTGGAAAGCAGCGCACTTGA	1598
Db	1321	GAAAGAGATCCGAGAGCGCGAGGCGCGCAAAACGGGAGCTCGAAACGGCGCAACGCTTGA	1380
Qy	1539	ATGGGAAAGGAAACCGGAGCAGGAACTCCTGAAATAGAGGAAACAAGGACAGGAGGAGAC	1658
Db	1381	ATGGGAAAGGAAACCGGAGCAGGAACTCCTGAAATAGAGGAAACAAGGACAGGAGGAT	1440
Qy	1659	CGTGCTCTGAAAGGCAAGAGAGAGAACTCTGGAGTTTGAGTTAGAGCTCTGAATGACA	1718
Db	1441	CGTGCTCTGAAAGGCAAGAGAGAGAACTCTGGAGTTGAATTAAGAGCTCTGAATGACA	1500
Qy	1719	AAAGCATCAGCTAGAAAGGAAAACTTCAGAGATTCAGGTGTGCACTGGCAACCCAGAGCA	1778
Db	1501	AAAGCATCAGCTAGAAAGGAAAACTTCAGAGATTCAGGTGTGCACTGGCAACCCAGAGCA	1560
Qy	1779	AGAAATTCAGAGCAGAAACAAGTCTAGAGAGCTGAAGTTGCTGAATCACCCACTTCA	1838
Db	1561	AGAAATTCAGAGCAGAAACAAGTCTAGAGAGCTGAAGTTGCGGAATCACCCACTTCA	1620
Qy	1839	GCAGCAGTTGCAAGAAATCTCAGCAAACTCTTGGAAGACTTATTCAGAGAAACAGATACT	1898
Db	1621	GCAACAGTTGCAAGAAATCTCAGCAAACTCTTGGAAGACTTATTCAGAGAAACAGATACT	1680
Qy	1899	CAGTGACCAAGTTAAAAACAAGTCCAGCAGAAACAGTTGCATAGAGACTCGCTTTTACCT	1958
Db	1681	CAGTGACCAAGTTAAAAACAAGTCCAGCAGAAACAGCTTGATAGAGATTCGCTTTTACCT	1740
Qy	1959	CAAAAGACCTTTGGAAGCAAGAGACCTGCCCGGAGAGAGCTCCGGGAGCAGCTGAGGA	2018
Db	1741	CAAAAGACCTTTGGAAGCAAGAGAACTGGCCGGAGAGAGCTTCCAGAGCAGCTGAGGA	1800
Qy	2019	GGTGAGAGAGAGACCAAGTCAAAAGCTCAGAGAGATTATGTTTCAACAACAGACTAA	2078
Db	1801	GGTGAGAGAGAGACCAAGTCAAAAGCTCAGAGAGATTATGTTTCAACAACAGACTAA	1860
Qy	2079	GGAACCTGAGAGATACATAGCAAAACAGCAATCCAGAGAGCAGAGTCCCTGAGGCGC	2138
Db	1861	GGAACCTGAGAGATACACAGTAAACAAGACTCAGAGAGCAGAGTCCATCGAGGCGGA	1920
Qy	2139	GCGATCTGAAGCAAGAAAGACAGAGAGAGAAAGCCTTGAGTTTAGAGAGCAAAAGAGAA	2198
Db	1921	GAGGCTTAAAGCAAGAAAGACAGAGAGAGAAAGCCTTGAGTTTGAGAGCAAAAGAGAA	1980
Qy	2199	CGCTCAGAGACCAAGTTCAAGAAAGGAGCAAGCAATGGCTGGAGCATGTGCAGCAGAGGA	2258
Db	1981	AGGTAGAGACCAAGTTCAAGAAAGGAGCAAGCAATGGAGAGCATGTGTCCAGCAGAGGA	2040
Qy	2259	GCAGCAGCGCCCCGAGAAACCCCAAGAGAGGACAGACTGAAGAGGAGAAACAGTTCAG	2318
Db	2041	GCAGCAGCGCCCTCGGAAACCCCAAGAGAGGACAAACTGAAGAGGAGAAACAGTTCAG	2100
Qy	2319	GAAAGAGAGCGCGGAAAGAGAGCCAAAGCCGGAATGCAGACAGCAGTCCGCTTTT	2378
Db	2101	GAAAGAGAGCGCGGAAAGAGAGCCAAAGCCGGAATGCAGACAGCAGAGTCCGCTTTT	2160

QY	2379	CCATTCGGATCAGAGCCAGCTTAAGCTGCGCCACCAGACACCTCGTCTTACCA	CAGAGAA	2438			
Db	2161	CCATTCCAATCAAGCCAGCTTAAGCCGG	---GCAGGACACCTCGGCCACGAGAGAA	2217			
QY	2439	AGGCCCGCTTACCAATTTCTGCA	CAGAGAGTGTAAAAGTGGTATATTAC	CAGAGCTGT	2498		
Db	2218	AGGTCCGTTTCAATCTCTGCA	CAGAGAGTGGCCAAAGTGGTGTATTAC	CGAGCGCTGT	2277		
QY	2499	CCCCCTTGAATCCAGAA	GTCAAGATCA	CAATCCAGCCAGAGATATTAGTCAT	---	2555	
Db	2278	CCCCCTTGAATCCAGAA	GTCAAGATCA	CAATCCAGCCAGAGATATTAGTCAT	GT	2337	
QY	2556	-----GGATGATGA	AAACCGAC	CTGAGAGACAGAGATGGCTGGAGAGAGCT		2603	
Db	2338	TAAAGGGGAATGGTGGATGA	AAACCGAC	CTGAGAGACAGAGATGGCTGGAGAGAGAC		2397	
QY	2604	GAAAGGGAGACGGGATGGTTC	CCCTGCAAACTAT	ATGCAAAAAATTC	CGAAAAATGAGT	2663	
Db	2398	GAAAGGGAGACAGAGATGGTTC	CCCTGCAAACTAT	ATGCAAAAAATTC	CGAAAAATGAGAT	2457	
QY	2664	TCCCACTTCAGACCAAA	CCAGTGAACCGATCTG	ACATCTGCCCCCTGCCCCCAAACTGGCTCT		2723	
Db	2458	TCCCACTTCAGACCAAA	CCAGTGAACCGATCTG	ACATCTGCCCCCTGCCCCCAAACTGGCTCT		2517	
QY	2724	GCGTGAGACCCCTGCTCT	CTTTGACAGTGA	CCCTCTTCTGAGCCCTCCAAACCCCAACAA		2783	
Db	2518	GCGTGAGACCCCTGCTCT	CTTTGACAGTGA	CCCTCTTCTGAGCCCTCCAAACCCCAACAA		2577	
QY	2784	CTGGGACAGCTTCA	AGTTCCACGTGGCC	CAGACGCAAAATGAGAA	CCAGAAACGGACAA	2843	
Db	2578	CTGGGACAGCTTCA	AGTTCCACGTGGCC	CAGACGCAAAATGAGAA	CCAGAAACGGACAA	2637	
QY	2844	CTGGGATTCGAGGGCGGGCT	CAGCCCTCTCTGAC	CCGTACCTGAGCTGAGCCAGTTACGGCA		2903	
Db	2638	CTGGGACAGCTGGGGCGGCT	CAGCCCTCTCTGAC	CCGTACCTGAGCTGAGCCAGTTACGGCA		2697	
QY	2904	GAGATCAGCCCTTTTAC	CCCCAGCCACAGCCACTG	AGCTCTCCCAATCTCCCGTCTTGCGCA		2963	
Db	2698	GAGGTCCGCTTTACAC	ACAGCCACAGCCACTG	AGCTCTCCCAATCTCCCGTCTTGCGCA		2757	
QY	2964	GGGTGAAAAGGTG	GAAGGGCTTCA	AAAGCGCAAAACCCCTGTATCCCTTGAGAGCGCAAAAAAGA		3023	
Db	2758	GGGTGAAAAGGTG	GAAGGGCTTCA	AAAGCGCAAAACCCCTGTATCCCTTGAGAGCGCAAAAAAGA		2817	
QY	3024	CAACCACTTAATTTTAA	CAAAAGTGA	CGTATCA	CCGTTCTTGAAACGACAGACATGTG	3083	
Db	2818	CAACCACTTAATTTTAA	CAAAAGTGA	CGTATCA	CCGTTCTTGAAACGACAGACATGTG	2877	
QY	3084	GTGGTTTGGAGAGTTCA	AGGTCA	GAAGGGTTGGTTC	CCCAAGTCTTACGTGAACCTCAT	3143	
Db	2878	GTGGTTTGGAGAGAGTTCA	AGGTCA	GAAGGGTTGGTTC	CCCAAGTCTTACGTGAACCTCAT	29937	
QY	3144	TTCAAGGGCCGTTA	AGAAATCA	CAAGAGATGATAC	TGCGCCCTTACCTGCTAG	3203	
Db	2998	TTCAAGGGCCGTTA	AGAAATCA	CAAGAGATGATAC	TGCGCCCTTACCTGCTAG	29997	
QY	3204	TCTAAAGAGATGGCTTC	CCCGGCGCCCAAGCC	ACCAATTC	CCGAGAGAGATTATTC	3263	
Db	2998	TCTAAAGAGATGGCTTC	CCCGGCGCCCAAGCC	ACCAATTC	CCGAGAGAGATTATTC	3057	
QY	3264	CATGTATCA	CAATCGAGAGTTCTG	AGCAAGAGATTAA	CTTTACGCAAGGGAGTGTGAT	3323	
Db	3058	TATGTATCA	CTTACAGAGTTCTG	AGCAATGAGATTAA	CTTTACGCAAGGGAGTGTGAT	3117	
QY	3324	TGTGATTAC	CAAGAAAGATG	TGATCGTGGAGCCGGAAC	CGGTGGCGCA	CAAGTCCGAGT	3383
Db	3118	TGTGATTAC	CAAGAAAGATG	TGATCGTGGAGCCGGAAC	CGGTGGCGCA	CAAGTCCGAGT	3177
QY	3384	CTTCCCTTTTAA	CTATGTGAGGCTTTAA	AGATTC	AGAGGGCTCTGGAAC	TGCTGGGAAAC	3443
Db	3178	CTTCCCTTTTAA	CTATGTGAGGCTTTAA	AGATTC	AGAGGGCTCTGGAAC	TGCTGGGAAAC	3237

QY 3444 AGGAGTTTGAAGAAAAAAGCTGAAATGTCAGGATTTGCTCTACGCTGCTACTG 3503
 Db 3238 AGGAGTTTGAAGAAAAAAGCTGAAATGTCAGGATTTGCTCTACGCTGCTACTG 3297
 QY 3504 TCCGGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3563
 Db 3298 TCCGGAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3357
 QY 3564 TGGATGTTGGAAGAGAACTGCAAGTTCGAGGAAAAAGCCAGATGAGGTTTCC 3623
 Db 3358 TGGATGTTGGAAGAGAACTGCAAGTTCGAGGAAAAAGCCAGATGAGGTTTCC 3417
 QY 3624 AGCAAAATTTATGCAAACTTTTAAAGCCCGGAAACAAATCAACCCCACTGAGCTAC 3683
 Db 3418 AGCAAAATTTATGCAAACTTTTAAAGCCCGGAAACAAATCAACCCCACTGAGCTAC 3477
 QY 3684 CAGACCGGAGTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3743
 Db 3478 CAGACCGGAGTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3537
 QY 3744 GAAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3803
 Db 3538 GAAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3597
 QY 3804 CCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3863
 Db 3598 CCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3657
 QY 3864 AAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3923
 Db 3658 AAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3716
 QY 3924 TCAGGCTTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3983
 Db 3717 TCAGGCTTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3776
 QY 3984 GGAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4043
 Db 3777 GGAAGTGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3836
 QY 4044 GAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4103
 Db 3837 GAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3896
 QY 4104 AGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4163
 Db 3897 AGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3955
 QY 4164 GGGGCTTTCCTAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4201
 Db 3956 -GTGCTTTCCTAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3992
 RESULT 5
 AF064243 5287_bp mRNA linear PRI 21-NOV-1998
 LOCUS AF064243 Homo sapiens intersecin short form mRNA, complete cds.
 DEFINITION AF064243
 ACCESSION AF064243.1 GI:3859852
 VERSION
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 5287)
 AUTHORS Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
 Antonarakis, S.E.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CML, 1 rue
 Michel-Servet, Geneva 4 CH-1211, Switzerland
 FEATURES
 source
 1..5287
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="21"
 /map="21q22.1-q22.2"
 /tissue_type="brain"
 /dev_stage="fetus"
 107..3769
 /codon_start=1
 /product="intersecin short form"
 /protein_id="AAC78610.1"
 /db_xref="GI:3859853"
 /translation="MAORPPRGSLDIWALTVEBRARHDOQFHSKXISGTTGDOA
 RNFPSGLPQVLAQIHALDMNDGMDVBERIAKLILKLOQYQLSALPVM
 KOQPAISAPPPFMGIAIMPPLTAVAPVWGSIPIVGMSPPLVSPVPAVPAVPA
 GAPVIOPLPAFPAFPAATLPKSSPSRSPGSOINTLQKAQSFVAVSPVAEWA
 OSSRLKRYQLFNSHDKTMGHLTGPOARTIMOSLPQALASIWNLSDIDODGLTA
 EEFILAMHLIDVAMSGOPLPVLPEPVIPEPFRVRSSEGI SVISSTVDORLPPEPV
 LEDBOOLEKTLPTVPEPDKRENERPERGVLLEKROALLBOORKEBLAOLERAQ
 EKERROERKROLELEKOLEKORLEERBEERKETERBEAKRELEKORLEME
 RNRROELNORNKQEDIVLAKKRTLELEALNDKRLQLEGLQDITRLLTQRO
 EIESNKRERLEALITHLQOLOSQOMLRLPEKQILNDOLKOVOONSILHRDLSV
 TLKRALFAKELARQHLRDOLDEVEKESKQLEIDIPNNOLKELEHNNKQLOKRS
 MEAERLKOKEBERKILEKOKERORAROWLEHMOEDDEHOKPRLHEBEKX
 KREBSVKKDEBEKROKADOKRLRLFHQDEPAKPAVOAHPSTDEKPLITSAQEN
 KVVYTRALYPRESNHDITTOPEGLIVAKGEMWDSOTGSPGLGGLKGTGWFPA
 NYAEKIPENEPVAPKPTDSTSAAPALALRETPAPLAVTSSESTPNNWABEST
 WPTSTNEKPEPNDAMAAQPSLTPVPSAGLORSAFTPATVTSSESTPNNWABEST
 GLQOALYPMRAKKNLHFNKNDVITVLEQODMMWFGEVQOGKMPKSVYKLSGP
 IRKTSMSGSESPASLKRVAAPAKVSGEELIAYTESSEQGLTQOQGVIL
 VTKDQMWNTGTVDKAGVPSNVYRLKDSGSGTAGTSGLSKKPEIAQVATSTAT
 GPBULTAPGQILIRKKNPGGMBEGLQARCKRQIGMPFANTVKLNPSTGSKTPT
 EEPSTALAAVCQVITGMVDTYANDDELAFNKGQI INVLNKEDPMWKGENVQGLP
 PSNVYKLTDDPSSQ"
 167..406
 /note="encodes EH domain"
 misc_feature
 767..1936
 /note="encodes EH domain"
 misc_feature
 2324..2524
 /note="encodes SH3 domain"
 misc_feature
 2843..3019
 /note="encodes SH3 domain"
 misc_feature
 3110..3286
 /note="encodes SH3 domain"
 misc_feature
 3326..3520
 /note="encodes SH3 domain"
 misc_feature
 3569..3748
 /note="encodes SH3 domain"
 BASE COUNT 1635 a 1169 c 1227 g 1256 t
 ORIGIN
 Query Match 67.7%; Score 3440.2; DB 9; Length 5287;
 Best Local Similarity 83.8%; Pred. No. 0;
 Matches 4154; Conservative 0; Mismatches 743; Indels 58; Gaps 21.
 QY 155 GCGGCTGCGGCTGCACTGATTTGTGTGAGGCGGCGCGGCAACCCGCGAGATG 224
 Db 13 GCGGCTGCGGCTGCACTGATTTGTGTGAGGCGGCGCGGCAACCCGCGAGATG 72
 QY 225 AGGCGTCATGAGCAAGGTGAACGAATAGAACATGAGTCTAGTTCCACACCTTCGG 284
 Db 73 AGGCGTCATGAGCAAGGTGAACGAATAGAACATGAGTCTAGTTCCACACCTTCGG 132
 QY 285 TGTAGCTGATGTCGGCCATTAAGTGTGAGGAGAAAGGCGCAAGATGACGACGATT 344
 Db 133 TGGAGCTGATGTCGGCCATTAAGTGTGAGGAGAAAGGCGCAAGATGATGACGACGATT 192

345 CTTAGCCGTAAGCCGATAGGGGATTTATTAAGTGATCAAGCAGAGAACTTTTTTT 404
193 CCAATGTTTAAAGCCAAATATCTGAATCAATTAATCTGATATCAAGTAAAGAACTTTTTTT 252
405 CCAATCTGGGTTTACCTCAGCCCTGTCTTAGCACAAATATGCGCGTAGCGGACATGATTA 464
253 TCAATCTGGGTTTACCTCAGCCCTGTCTTAGCAAGATATGGGCACTAGCTGACATGATTA 312
465 CGATGGAAGATGATCAAGTGGAAATTTTCCATAGCCATGAACTTATCAACTGAACT 524
313 TGAATGGAAGATGATCAAGTGGAAATTTTCCATAGCCATGAACTTATCAACTGAACT 372
525 ACAAGATATCAGCTCCCTCCACACTTCCCTGTCAAGAAAGCAACAGAGGCTAT 584
373 ACAAGATATCAGCTCCCTCCACACTTCCCTGTCAAGAAAGCAACAGAGGCTAT 432
585 TTCCAGTCCAGCAGATTTGTATAGAGGATTTGCTAGCATGCCCACCTACAGCTGT 644
433 TTCTAGGCGCACACCATTTGGTATGGAGGATAGCGCAGCATGCCACCGCTTAGAGCTGT 492
645 TGCTCTGTGCGCAATGGGCTCCATTCAGTTGTTGGAATGTCTCCACCTTAGTATCTTC 704
493 TGCTCTGAGTCCAAATGGGATCCATTCAGTTGTTGGAATGTCTCCACCTTAGTATCTTC 552
705 TGTCCTCAGCAGAGAGTGGCCCTCCCTGCTTAAAGGGGCTCCCTCCCTATACAGCCCT 764
553 TGTTCCACAGAGCTGTGCCCCCTGCTTAAAGGGGCTCCCTCCCTGTTATACAACTCT 612
765 GCTGCGTTTGGCGATCTCTGAGCCACATAGCCAAAGATTTCTTCCAGCAGATCTG 824
613 GCTGCGATTTGGCTATCTCGAGGCAATGCGCAAGGTTCTTCTTAGAGATCTG 672
825 TCCAGGTCACAAATTAACAATAAGTACAGAGGCAATATTCATGATGCGCAGCGC 884
673 TCCAGGTCACAAATTAACAATAAGTACAGAGGCAATATTCATGATGCGCAGCGT 732
885 CCCTCCAGCAGAGAAATGGGCTGTGCTCAGTCATCAAGGCTGAAATCAGGAGTTAT 944
733 CCCTCCAGGTCAGAGTGGGCTGTGCTCAGTCATCAAGGCTGAAATCAGGAGTTAT 792
945 CAACAGCCACGACAAATATAGATGAGCACTTAAACAGGTCGCCAGGCAAGAACTATCT 1004
793 CAATAGTCATGACAAATATAGATGAGCACTTAAACAGGTCGCCAGGCAAGAACTATCT 852
1005 CATGCAATCAAGTTATCCCGAGGCTCAGCTGCTTCAATATGGAATCTTTCTGACATTGA 1064
853 TATGCAATCAAGTTATCCCGAGGCTCAGCTGCTTCAATATGGAATCTTTCTGACATTGA 912
1065 TCAAGATGGAATACTCAGTCGAGAGAAATTTATCTACTATGACCTTAATGATGTGC 1124
913 TCAAGATGGAATACTCAGTCGAGAGAAATTTATCTGCAATGACCTCATATGATGATG 972
1125 CATGCTGTGTCAGCCACTGCGGCCCTGCTCTGCTCAGAAATACATCCCTCTCTTCTGAG 1184
973 TATGCTGTGTCAGCCACTGCGGCCCTGCTCTGCTCAGAAATACATCCCTCTTCTTAG 1032
1185 AAGAGTTGCTCTCGGCAATGGGATTCGCTCATTAAGCTCTTCTTGTGATCAGAGCT 1244
1033 AAGAGTTGCTCTCGGCAATGGGATTCGCTCATTAAGCTCTTCTTGTGATCAGAGCT 1092
1245 GCCTGAGAGAGCCGTGTGAGAGATGAGCAGAGC---CAGAGAGAAATGCGCTGTAC 1301
1093 ACCAGAGAGAGCCGTTTATGAGAGATGAAACAAACAAATTTAGAAAGAAATTAACCTGTAC 1152
1302 ATTGGAATGATGAAGACGCGAGAACTTCAGCAGAGCAGTGTGAGCTGAGAAAGCGCG 1361
1153 GTTTGAAGATTAAGAAAGCGGAGAACTTTGAAGTGTGCAACTGTGAACCTGGAGAAAG 1212
1362 CCAAGCGCTCTTTGAGCAGCAGCGCAAGAGAGAGAGCGGTTGGCTCAGCTGAGAGCGGC 1421
1213 GCAAGCTCTCTCGAAGCAGCAGCGCAAGAGAGAGAGCGCTGCGCCAGCTGAGAGCGGC 1272

1422 CGACAGAGAGAGAAAGCGCGAGCGCCAGAGAGCAGAGAGCGCAAGCGGCTGAGACT 1481
1273 GAGACAGAGAGAGAGAGAGCGCTGAGCGCCAGAGAGCAGAGAGCGCAAGCACTGAGACT 1332
1482 GAGAGACAGCTGAGAGAGCAGCGGAGCTGAGCGGCAAGAGAGAGAGAGAGAGAGAG 1541
1333 GAGAGACAGCTGAGAGAGCAGCGGAGCTGAGCGGCAAGAGAGAGAGAGAGAGAGAG 1392
1542 GAGATGAGAGAGCGCGAGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1601
1393 AGAATTTGAGAGCGAGAGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1452
1602 GGAAG 1661
1453 GGAAG 1512
1662 GGTCTGAG 1721
1513 TGTACTGAG 1572
1722 GCATCAGCTAG 1781
1573 GCATCAACTAG 1632
1782 AATTGAG 1841
1633 AATTGAG 1692
1842 GCAATGAG 1901
1693 ACAATTCAG 1752
1902 TGACCACTGAG 1961
1753 TGACCACTGAG 1812
1962 AAGAGCTTGAAG 2021
1813 AAGAGCTTGAAG 1872
2022 GAG 2081
1873 GAG 1932
2082 ACTGAGAGAGATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2141
1933 ACTGAGAGAGATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1992
2142 ACTGAG 2201
1993 ACTGAG 2052
2202 TCAG 2258
2053 CCAAG 2112
2259 GCAG 2318
2113 GCATCAG 2172
2319 GAG 2378
2173 AAG 2232
2379 CCAATCGCATCAG 2438
2233 CCAATCAACAG 2292
2439 AGGCGCGCTTACATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2498
2293 AGGTCACATTAACATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2352
2499 CCCCTTTGAATCCAG 2555

Db 2353 CCCCTTGAATCAGAAAGCCATGATGTAATCACTATCCAGCCAGAGACATGTATGTGT 2412
 Qy 2556 -----GCTGATGAAAAGCCAGACTGAGAGCCAGATGTGCTTGGAGAGAGCT 2603
 Db 2413 TAAAGGGGAATGGGTGATGAAAAGCCAACTGGAGAACCCGGCTGGCTTGGAGAGAAAT 2472
 Qy 2604 GAAAGGGAAGACGGGATGTGCTCCCTGCAAACTATGCAAGAAAAGATTCCAGAAAATGAGGT 2663
 Db 2473 AAAAGGAAAGACAGGGTGTGCTTCCCTGCAAACTATGCAAGAAAATGAGGT 2532
 Qy 2664 TCCCATCCAGCCAAACAGATGACCGATCTGACATCTGCCCCCTGCCCAACTGGCTCT 2723
 Db 2533 TCCCTCTCAGTGAAGACAGTACTGATTCACATCTGCCCTGCCCAACTGGCTCT 2592
 Qy 2724 GCGTGAAGACCCCTGCTCTTCTTGGCAATGACCTTCTTGAAGCTCTCCACAACTCCCAACA 2783
 Db 2593 GCGTGAAGACCCCTGCTCTTGGCAATGACCTTCTTGAAGCTCTCCACAACTCCCAACA 2652
 Qy 2784 CTGGGACACTTCAGTTCCAGCTGGCCGACGACTCAACGAGAGCCAGAAAAGGAGCA 2843
 Db 2653 CTGGGACCACTTCAGTTCCAGCTGGCCGACGACGAGATGAGAAACGAGAAAGGATTA 2712
 Qy 2844 CTGGGATAGTGGGCGGCTCAGCCTTCTCTGACCGTACTAGTGGCCAGTTACGAGCA 2903
 Db 2713 CTGGGATGATGGGAGCCGACGCTCTCTCACCCTTCCAGTGGCCGACGTTAAGCA 2772
 Qy 2904 GAGATCAGCTTTACCCGACCAACGACCTGCTCTCCCTCCCACTCCCGTCTGGGCA 2963
 Db 2773 GAGGTCGCGCTTACTCCAGCCACGCGCACTGGCTCTCCCGTCTCTGCTAGGCCA 2832
 Qy 2964 GGGTGAAGAGTGAAGGGCTCAAGCGGAGCGGCTGATCCCTGAGAGCCAAAAAGA 3023
 Db 2833 GGGTGAAGAGTGAAGGGCTCAAGCGGCTGATCCCTGAGAGCCAAAAAGA 2892
 Qy 3024 CAACCACTTAATTTTAAACAAAGTACGTCATCAACGTTCTGGAACAGCAAGACATGTG 3083
 Db 2893 CAACCACTTAATTTTAAACAAAGTACGTCATCAACGTTCTGGAACAGCAAGACATGTG 2952
 Qy 3084 GTGTTTGAAGAACTTCAAGGTCAGAGGGTGTGCTCCCAAGTCTTACGTAAGTCAAT 3143
 Db 2953 GTGTTTGAAGAACTTCAAGGTCAGAGGGTGTGCTCCCAAGTCTTACGTAAGTCAAT 3012
 Qy 3144 TTCAGGGCCCTTAAGGAATTCAGACATGATCTGGCCCTTACGTAAGTCTGCTAG 3203
 Db 3013 TTCAGGGCCCTTAAGGAATTCAGACATGATCTGGCTTCTACAGAGTCTGCTAG 3072
 Qy 3204 TCTAAAGAGTGTGCTTCCCGGCGCCAGCCAGCCATTCGCGAGAGAGATTATTC 3263
 Db 3073 TCTAAAGAGTGTGCTTCCCGGCGCCAGCCAGCCAGTCTTGGAGAGAGATTATTC 3132
 Qy 3264 CATGTACATACGAGAGTTCGAGCAGAGAGATTAACTCTTTCAGCAAGGGAGTGTAT 3323
 Db 3133 CATGTACATACGAGAGTTCGAGCAGAGAGATTAACTCTTTCAGCAAGGGAGTGTAT 3192
 Qy 3324 TGTGTTTCAAGAAAGATGTGACGTTGTGACGAGGAGGAGGTGGGAGAAATTCGGAGT 3383
 Db 3193 TGTGTTTCAAGAAAGATGTGACGTTGTGACGAGGAGGAGGTGGGAGAAATTCGGAGT 3252
 Qy 3384 CTTCCTTTCTAATCTGTGAGCTTAAAGATTCAAGAGGCTCTGGAACCTGTGGGAAAAC 3443
 Db 3253 CTTCCTTTCTAATCTGTGAGCTTAAAGATTCAAGAGGCTCTGGAACCTGTGGGAAAAC 3312
 Qy 3444 AGGAGATTAGGAAAAAACTGAAATTGCCAGGTTATGCTTCTCTACGCTTACTGG 3503
 Db 3313 AGGAGATTAGGAAAAAACTGAAATTGCCAGGTTATGCTTCTCTACGCTTACTGG 3372
 Qy 3504 TCCCGAACAACCTCAACCTGGCTCTGGGAGGCTGAGTTCGATCCGGAATAAACCAG 3563
 Db 3373 CCCCAGACGCTCACTCTGCCCCCTGGTACGATTTGATCCGAAAAAGAACCCAG 3432
 Qy 3564 TGGATGTGGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCAGATAGGGTGTGTTCC 3623

Db 3433 TGGATGTGGAGAGAGCTCAGACAGTGGAAAAAGCCAGATAGCTGTGTTCC 3492
 Qy 3624 AGCAATTAATGTCAACTTTCTAAGCCCGGGAACAAAGCAAAATCACCCTCACTGACTAC 3683
 Db 3493 AGCTAATTAATGTAAGCTTTCTAAGCCCTGGGACGAGCAAAATCACTCCAAAGAGCAC 3552
 Qy 3684 CAAGACCGAGTGCAGCAGAGCTGTGCAAGGTATCGGAGATGTAAGATTAACCGGCA 3743
 Db 3553 TAAAGTCAACAGATTAAGCGGACAGTGTGCAAGGTATGAGATGTAAGATTAACCGGCA 3612
 Qy 3744 GAACGATGACGAATAGCTTCAAGAAAGCCAGATCAATCAAGCTTCAACAGAGAG 3803
 Db 3613 GAATGACGATAGCTGCTTCAACAGAGCCAGATCAATCAAGCTTCAACAGAGAG 3672
 Qy 3804 CCGGAGTGTGGAAAAAGAGAGTCAAGTGGGAGTGTGGGCTTCCATCAATTAAT 3863
 Db 3673 CCTCAGTGTGGAAAAAGAGAGTCAAGTGGGAGTGTGGGCTTCCATCAATTAAT 3732
 Qy 3864 AAAGCTGACACAGACATGAGACCCGACGAGCAATGAATCATATGTTGCCATCCCGCC 3923
 Db 3733 GAAGCTGACACAGACATGAGACCCGACGAGCAATGAATCATATGTTGCCATCCCGCC 3792
 Qy 3924 TCAGCTTGAAGAGTCTCAAGAGAGCCAGATATCCATATCACTGCGCAGAGAGATAG 3983
 Db 3793 TCAGCTTGAAGAGTCTCAAGAGAGCCAGATATCCATATCACTGCGCAGAGAGATAG 3852
 Qy 3984 GAGATGCAAGCTTATCATGTGACTTGCAGATATCATCTTACCTTCTGAGTAGAA 4043
 Db 3853 GAGATGCAAGCTTATCATGTGACTTGCAGATATCATCTTACCTTCTGAGTAGAA 3912
 Qy 4044 GAACCTCAGTGCAGAGCAGTTTACCTTATGACCTTATGATGATGCAAGATGCTG 4103
 Db 3913 GAACCTCAGTGCAGAGCAGTTTACCTTATGACCTTATGATGATGCAAGATGCTG 3972
 Qy 4104 AGTCACTGCTGACAGAGCAGAGCA--AATTCAGAACTGACAGAGGTGTGGTCTT 4161
 Db 3973 AGTATTAATCTGACAGATAGAGCAAAATTAACAAACACAGGGATGTGGTCTT 4032
 Qy 4162 TTGGGGCTTCTTACTGACTCAGACTGAC--CGGCCCGCTTACACAGGGCGCTTCAAT 4220
 Db 4033 TTGGGGCTTCTTACTGACTCAGACTGAC--CGGCCCGCTTACACAGGGCGCTTCAAT 4092
 Qy 4221 A-GTTTAAAGATTAATTTTAAATGTATTTAGCTTTAATAAATCTCAATCAAT 4279
 Db 4093 AGTITTAATTAATTTTAAATGTATTTAGCTTTAATAAATCTCAATCAAT 4152
 Qy 4280 ACTTCTTGCCTATTTGTTTAAACAAAGCAACCACTATCAAGAGTGTGCTGCGG 4339
 Db 4153 GACTTCTTGTGATTTTGTGTTTGAAGAAAGCCAGCTATCAAGGAATG-----CTGAT 4207
 Qy 4340 ACGATTAATGCTGTTCGCGGCGTACCGTAACTGAGAGCTGCTGTAACCTTGCCGT 4399
 Db 4208 GTGCTATTAATAATTTGTTCAAAATGTCCATTAATCTGAGACTGTATGTTTTCATTT 4267
 Qy 4400 TGTCAAGTTCCTCAACACAACTGTGTAGTTGGGGCGTGTCC-----TGCCGTAGAGA 4455
 Db 4268 TGTCAAGTTCCTCAACACAACTGTGTAGTTGGGGCGTGTGTTCCTTACATAGAGG 4327
 Qy 4456 CAGAGAGATGGGTGTAACCTGTTTGAAGATGTATGATAGACTGAGCTGCTATGGA 4515
 Db 4328 CAGAGAGATGTAGTATCTGTTTAAAGAGATTA--GAATGAGCCCAATTAAGAGAG 4386
 Qy 4516 GGGGTTAGCTTGTGTGACCATCAAGTATAC--TGTGCGCATGTACCATCTGTACCG 4574
 Db 4387 GTGTTTGTGCTTGTGTGTATGATGCTGTACCTGTTTGAAGCATGATATCATCTGTA 4446
 Qy 4575 AAGAGTACCT--CTTCCCTCAAGCTTAAACCAACCAAGCTGTACAGTGTCTCACTACT 4633
 Db 4447 CATTAAGATTAATGTTCTTCAATGCAAGGCTATTAACCTTGTACATGATCTCAATCAT 4506
 Qy 4634 --GCATTAATTTTATTTGACAGTGAACCTTGTAGCCACTGAGAGAACAC--CCATGTTT 4690
 Db 4507 TGCATTAATTTTATTTTGAACAGTGAACCTTGTAGCCACTGAGAGAACACCTGTGTTT 4566

Db 721 GTTATTCACAGTCACAGACAGACCATGATGAGCATTTAACAGTCCCGAGGACAGAAC 780
QY TATTCATGACATCAAGTTTATCCAGAGTCAGCTGGCTTCAATATGAAATCTTTCTGA 1058
Db 999 TATTCATGACATCAAGTTTATCCAGAGTCAGCTGGCTTCAATATGAAATCTTTCTGA 1058
Db 781 TATTCATGACATCAAGTTTATCCAGAGTCAGCTGGCTTCAATATGAAATCTTTCTGA 840
QY 1059 CATTTGCAAGATGAGAAATCTCACTGACAGAAATTTATCTGATGACACTTAATTA 1118
Db 841 CATTTGCAAGATGAGAAATCTCACTGACAGAAATTTATCTGATGACACTTAATTA 900
QY 1119 TGTTCGATGTCGTGTGACCACTGCGCCCTCTGCTCCAGAAATACATCCCTCTTC 1178
Db 901 TGTTCGATGTCGTGTGACCACTGCGCCCTCTGCTCCAGAAATACATCCCTCTTC 960
QY 1179 CTTCAGAAAGTTCGCTCCGCGAGTGGATGTCCTGCTTAAAGTCTTCTTCTGTGATCA 1238
Db 961 CTTCAGAAAGTTCGCTCCGCGAGTGGATGTCCTGCTTAAAGTCTTCTTCTGTGATCA 1020
QY 1239 GAGGCTGCTGAGAGCCGCTGTCAGAGATGAGACAGCCAGAGAAATGCTGCTGT 1298
Db 1021 GGGGCTGCCCGAGGACCATCTGACAGATGAGACAGAGTGGAAAGAAAGCTGCTGT 1080
QY 1299 GACATTTGAAGATTAAGAGCGGAGAACTTTCAGCGAGCGAGTGTGAGCTGAGAGCG 1358
Db 1081 GACATTTGAAGATTAAGAGCGGAGAACTTTCAGCGAGCGAACCTGAGCTGAGAGCG 1140
QY 1359 CGGCCAAGGCTCTTTGAGACAGAGCGCAAGAGACAGAGCGGTTGGCTCACTGAGAGG 1418
Db 1141 CAGGCGAGGCTCTCTGAGACAGAGCGCAAGAGAGCGAGCGCTGCTCACTGAGAGG 1200
QY 1419 CGCCGAGCAGAGAGAGAGAGCGGAGCGCCAGAGCAGAGAGCGCAAGCGAGCTGGA 1478
Db 1201 CGCCGAGCAGAGAGAGAGAGCGGAGCGCCAGAGCAGAGAGCGCAAGCGAGCTGGA 1260
QY 1479 GCTGAGAAAGCAGCTGAGAAAGCAGCGAGACTGAGCGCAGCGAGAGAGAGAGAG 1538
Db 1261 GCTGAGAAAGCAGCTGAGAAAGCAGCGAGACTGAGAGGCAAGAGAGAGAGAGAG 1320
QY 1539 GAAAGGATTCAGAGAGCGCGAGCGCCAGAACTGGAAAGCGCACTGGAATCTTGA 1598
Db 1321 GAAAGGATTCAGAGAGCGCGAGCGCCAGAACTGGAAAGCGCACTGGAATCTTGA 1380
QY 1599 ATGGAACCGAAACCGAGACAGAGAACTCTGAATTCAGAGAAACAAGAGAGAGAGGAC 1658
Db 1381 ATGGAACCGAAACCGAGACAGAGAACTCTGAATTCAGAGAAACAAGAGAGAGGAC 1440
QY 1659 CGTGTCTTGAAGGCAAGAGAGAACTCTGAATTCAGAGAAACAAGAGAGAGGAC 1718
Db 1441 CGTGTCTTGAAGGCAAGAGAGAACTCTGAATTCAGAGAAACAAGAGAGAGGAC 1500
QY 1719 AAAAGCATCAGCTTGAAGAGAAATTTCAAGATTCAGAGTTCGACTGGCAACCGAGGCA 1778
Db 1501 AAAAGCATCAGCTTGAAGAGAAATTTCAAGATTCAGAGTTCGACTGGCAACCGAGGCA 1560
QY 1779 AGAAATTTGAGAGCAGCAAGCTCTAGAGACTTAAGATTTGCGAATCAACCCACTTAA 1838
Db 1561 AGAAATTTGAGAGCAGCAAGCTCTAGAGACTTAAGATTTGCGAATCAACCCACTTAA 1620
QY 1839 GCAGCAGTTGACAGAACTTCAGCAAAATGCTTGGAAAGCTTAATTCAGAGAAACAGATCT 1898
Db 1621 GCAGCAGTTGACAGAACTTCAGCAAAATGCTTGGAAAGCTTAATTCAGAGAAACAGATCT 1680
QY 1899 CAGTGAACAGTTTAAACAGTCCAGAGAAACAGTTCATAGAGACTCGCTTCTTACCT 1958
Db 1681 CAGTGAACAGTTTAAACAGTCCAGAGAAACAGTTCATAGAGACTCGCTTCTTACCT 1740
QY 1959 CAAAAGAGCTTGAAGCAAAAGAGCTGGCCCGGACAGCTCCGGAGAGAGCTGAGCA 2018
Db 1741 CAAAAGAGCTTGAAGCAAAAGAGCTGGCCCGGACAGCTTCCGAGAGAGCTGAGCA 1800
QY 2019 GGTGAGAGAGAGACAGGTCAAAGCTGACAGAGATTTGATTTTCAACAACAGCTGAA 2078
Db 1801 GGTGAGAGAGAGACAGGTCAAAGCTGACAGAGATTTGATTTTCAACAACAGCTGAA 1860

QY 2079 GGAACGTGAGAGATTAACATAGCAACAGCAATCCAGAGAGAGAGTCCCTGGAGCAGC 2138
Db 1861 GGAACGTGAGAGATTAACATAGCAACAGCAATCCAGAGAGAGAGTCCATGAGCGGA 1920
QY 2139 GCGACTGAAGCAGAAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
Db 1921 GAGGCTTAAGCAGAAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 2199 CGCTCAGAGACAGATTCAGAGAAAGGACAGCAATGCTGAGCATGTGACAGAGAGA 2258
Db 1981 AGGTGAGAGACAGATTCAGAGAAAGGACAGCAATGCTGAGCATGTGACAGAGAGA 2040
QY 2259 GAGGCGAGCGCCCGGAGAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2318
Db 2041 GAGAGAGCGCTTCGAGAAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2319 GAAAGAGAGAGCGAG 2378
Db 2101 GAAAGAGAGAGCGAG 2160
QY 2379 CCATCCGATCAAGAGCCAGCTTAAGCTGAGCAACCAAGGCACTTGGTTCACAGAGAA 2438
Db 2161 CCATCCGATCAAGAGCCAGCTTAAGCTGAGCAACCAAGGCACTTGGTTCACAGAGAA 2217
QY 2439 AGGCGCGCTTACATTTCTGCAACAGAGAGTAAAGAGTAAATTAACCGAGCGCTGA 2498
Db 2218 AGGCGCGCTTACATTTCTGCAACAGAGAGTAAAGAGTAAATTAACCGAGCGCTGA 2277
QY 2499 CCCCTTTGAATCAGAAAGTCAAGATGATCAATCCAGAGAGAGAGATATGATCAT --- 2555
Db 2278 CCCCTTTGAATCAGAAAGTCAAGATGATCAATCCAGAGAGAGATATGATGATGAT 2337
QY 2556 -----GCTGATGAAGCCAGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGT 2603
Db 2338 TAAAGGAGAGATGGGTGATGAAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2397
QY 2604 GAAAGGAGAGAGAGAGATGGTTCCTGCAAACTATCAGAAAGAGATTCAGAAAGATGAGT 2663
Db 2398 GAAAGGAGAGAGAGAGATGGTTCCTGCAAACTATCAGAAAGAGATTCAGAAAGATGAGT 2457
QY 2664 TCCCACTCAGAGCAAAACAGTACAGATCTGATCTGAGAGAGAGAGAGAGAGAGAG 2723
Db 2458 TCCCACTCAGAGCAAAACAGTACAGATCTGATCTGAGAGAGAGAGAGAGAGAGAG 2517
QY 2724 GGTGAGAGCCCTGCTCTTTGCAAGTGAACCTTTCTGAGAGAGAGAGAGAGAGAG 2783
Db 2518 GGTGAGAGCCCTGCTCTTTGCAAGTGAACCTTTCTGAGAGAGAGAGAGAGAGAG 2577
QY 2784 CTGGGCAAGCTTCAAGTTCAGAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2843
Db 2578 CTGGGCAAGCTTCAAGTTCAGAGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 2637
QY 2844 CTGGGATAGTGGGAGGCTCAGCTTCTCTGACCGATCTAGTGGGCACTTGAAGCA 2903
Db 2638 CTGGGATAGTGGGAGGCTCAGCTTCTCTGACCGATCTAGTGGGCACTTGAAGCA 2697
QY 2904 GAGATCAGCTTACCCAG 2963
Db 2698 GAGATCAGCTTACCCAG 2757
QY 2964 GGGTGAAGAGTGAAGAGGCTCAAGCGCAAGCCCTGTATCCCTGAGAGAGAGAGAGAG 3023
Db 2758 GGGTGAAGAGTGAAGAGGCTCAAGCGCAAGCCCTGTATCCCTGAGAGAGAGAGAGAG 2817
QY 3024 CAACCACTTAATTTTAAACAAAGTGAAGTCAACAGCTTGAAGAGAGAGAGAGAGAG 3083
Db 2818 CAACCACTTAATTTTAAACAAAGTGAAGTCAACAGCTTGAAGAGAGAGAGAGAGAG 2877
QY 3084 GTGGTTTGAAGAGTTCAGAGTTCAGAGAGAGTGTTCCTCCAGAGTTCAGAGAGAG 3143
Db 2878 GTGGTTTGAAGAGTTCAGAGTTCAGAGAGAGTGTTCCTCCAGAGTTCAGAGAGAG 2937

QY	3144	TTGAGGGCCCGTAAGAAATCCAGAGATGATCTGAGCCCTTAATGAAATGCTGCTAG	3203
Db	2938	TTGAGGGCCCGTAAGAAATCCAGAGATGATCTGAGCCCTTAAGAGCTCCTCTAG	2997
QY	3204	TTTAAAGAGAGTGGCTTCCCGGCGCCAGAGCCAGCACTTCCCGAGAGAATTAATTC	3263
Db	2998	TTTAAAGAGAGTGGCTTCCCGAGCGCCAGAGCCAGCACTTCCCTGAGAA-----	3046
QY	3284	CATGTACACATACGAGACTTGTAGCAAGAGATTTAACTTTACGACAAAGGAGTGTAT	3323
Db	3047	-----	3046
QY	3324	TGTGATTACAAAGAAAGATGGTGACTGTGTGACGGGAAAGGTGGGCGACAAGTCCGAGT	3383
Db	3047	-----	3046
QY	3394	CTTCCCTTCTAATATGTAGAGCTTAAAGATTCAAGGGCTCTGMACTGTGGGAAAAC	3443
Db	3047	-----	3046
QY	3444	AGGAGCTTAAAGAAAAAACTGAAATTTGCCAGTTATTTGCTTCTAAGCTGCTACTGG	3503
Db	3047	-----	3046
QY	3504	TCCCGAACACTCACCTTGCTCTCTGGGAGCTGATTTCTGATCCGGAAAAAGAACCCAG	3563
Db	3085	TCCTGAACAGCTCACCTTGCTCTCTGGGAGCTGATTTTGAATCCGAAAAAGAACCCGG	3144
QY	3554	TGGATGTGTGGGAAGAGAACTGCAAGCTCGAGGAAAAAGCGCCAGATAGGTGTTTC	3622
Db	3145	TGGATGTGTGGGAAGAGAACTCAAGCTCGAGGAAAAAGCGCCAGATAGGTGTTTC	3204
QY	3624	AGCAATATATGTCAACTTCTTAAGCCCGGAAACAAGCAAAATCACTCCAACTGAGCTAC	3683
Db	3205	AGCAATATATGTCAACTTCTTAAGCCCTGGAACACAGCAAAATCACTCCAACTGAGCTAC	3266
QY	3684	CAAGACCCGAGTGCAGCCAGCAGCAGTGTGCCAGGTGATCGGATGTACATTAACCCGCCA	3743
Db	3265	CAAGACCCGAGTGCAGCCAGCAGCAGTGTGCCAGGTGATCGGATGTACATTAACCCGCCA	3324
QY	3744	GAAAGATGACGAACTAAGCTTCAGAAAGCCAGATCATATCAACGTCTCAACAAGAGAGA	3803
Db	3325	GAAATACATAGACTAGCTTCAGCAAAAGCCAGATCATCAACGTCTTCAGAGAGAGA	3388
QY	3804	CCCGGACTGTGTGAAGAGGAAGTCAGTGGGCAAGTTGGGCTCTTCCATCCAAATTAATGT	3863
Db	3385	CCCGGACTGTGTGAAGAGGAAGTCAGTGGGCAAGTTGGGCTCTTCCATCCAAATTAATGT	3444
QY	3884	AAAGCTGACCAACAGACATGAGACCCAGCCAGACATGAATCATATTTGTTCATCCCCCC	3922
Db	3445	GAAAGTGCACCAAGACATGAGACCCAGCCAGACATGAATCATATTTGTTCAT-CCCCC	3503
QY	3924	TCAGGCTTGAAGGCTCTCAAAAGAGACCACTATCCCATATCACTGCCAGAGGATGTATG	3983
Db	3504	TCAGGCTTGAAGGCTCTCAAAAGAGACCACTATCCCATATCACTGCCAGAGGATGTATG	3566
QY	3984	GGAGATGCAGCTTGAATCATGTGACTTGTGACGATATATCATCTACTGCTTCTGATGAA	4043
Db	3564	GGAGATGCAGCTTGAATCATGTGACTTGTGACGATATATCATCTACTGCTTCTGATGAA	3623
QY	4044	GAACTCACATGACAGAGAGTTTACCTCATTTGACTTGAATGTGATTCGAAATGTGTG	4103
Db	3624	GAACTCACATGACAGAGAGTTTACCTCATTTGACTTGAATGTGATTCGAAATGTGTG	3688
QY	4104	AGTCACTGCTGACAGAGCAGAGCAAGCAAAATTGCAAACTGCAACAGGCTGCTGCTTT	4163
Db	3684	AGTCTCTGCTGACAGAGCAGAGCAAAATTGCAAACTGCAACAGGCTGCTGCTTT	3744
QY	4164	GGGCTTTCTTAGTCACTCAACTGACCGGCCCCGCT 4201	
Db	3743	GTGCTTTCTTAGTCACTCAACTGACCGGCCCCGCT 3779	

RESULT 7	AF114487	6439 bp	mRNA	linear	PRI 16-JUL-2002
LOCUS	AF114487				
DEFINITION	Homo sapiens interseectin long isoform (ITSN) mRNA, complete cds.				
ACCESSION	AF114487				
VERSION	AF114487.1	GI:4808822			
KEYWORDS					
ORGANISM	Homo sapiens (human)				
AUTHORS	Hom sapiens Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo. 1 (bases 1 to 6439)				
TITLE	Pucharcos, C., Fuentes, J.J., Caase, C., de la Luna, S., Alcantara, S., Arbones, M.L., Soriano, E., Estivill, X. and Pritchard, M. Alu-Splice cloning of human interseectin (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down syndrome Eur. J. Hum. Genet. 7 (6), 704-712 (1999)				
JOURNAL	Eur. J. Hum. Genet. 7 (6), 704-712 (1999)				
MEDLINE	99415290				
PUBMED	10482960				
REFERENCE	2 (bases 1 to 6439) Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X. Direct Submission				
AUTHORS	Submitted (16-DEC-1998) Medical and Molecular Genetics Center, Cancer Research Institute, L'Hospitalet de l'lo., Avia. Caetelldetels Km. 2,7, Barcelona 08907, Spain				
JOURNAL	Location/Qualifiers				
FEATURES	1..6439				
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /chromosome="21" /map="21q22.1-q22.2" 1..6439 /gene="ITSN" 269..5434 /gene="ITSN" /codon_start=1 /product="interseectin long isoform" /protein_id="A0D2952.1" /db_xref="GI:4808823" /translation="MAQPTTFPGSGDIATVYERAKHNOQFHSKLPIGSLFTGDOA RNFPGSLGPQVLAQIWMALANDNDRNQVPSRIAMKILKLGQYQIPALPPM KQGVALISSAPAFGMGSIAMPTLTAAVAPPMKSIIPVNGSPITLVSSVPPAAVPLAN GAPVLIQPLPAFAIPATLIPKSSFSFGSGSQNLTLQKQSGFDVASVPPVAMVAP QSSALVQPLFNSHDKTMSGHLNGPOARTLMQSSLPQALASIMNLSIDIDQGLTA EETLIQALIDVAMSGQPLPPVLPPEYIPSPFVRVSGSISVTSVQRLPBEV LEBQIQLLEKKLPTTFEDKRENRBNLEEKRRQALLLEQKQREQLAQLRABE RKSEKEROEQRQLELEKQLEKRENERLEERREKLEIRREAKRELEKQLEWE RNRQELINLRNKEQDITVILKAKKLTLEELANDKQGLGKQLDIRICRLTQKQ EISTEKNRELRILAEITHLQOQLOESQOMGRLIPKQILNDLQKQVNSLRHDSLV TLKRALFAKELARHLELDQDEVKEKERSGLQEDINNOQLREIRHNKQLOKQKS MEARLKQKOEORHITIELEKQKEAORABERKQMEIENHQOEDHORPKLHEEKL KRBSYVKKDGEKGEKQKADQKGLRHLHQEPAKRVQAPSTAEKGPITISAGENV KVYVYRVLVPESSHDEITIPQDILVWGVGWDVDSQTEPEMLGSELGKGMFPA NVYAEKIPENVEPAPVVKVEITTSAPAKALRETPAPLVAITSESPSTPPNMVADFSST WPTSTNEKPEPTMDWAAQNPSTLVPSAGQLRBPAFTPATIGSSPSPIIDGGEKVE GLOQALYPRKAKNDNINFNKNDVITLVEQODMMMGVEVQGGKGMFKSVYLISGP IRKTSMDSSSPSSPALIKRVSAPKAVSGEFTLAMYESSBOGDLTFQGGDVL VTYKDDGMWGTGVDKAGVPPSNVYRLKDSBEGSTAGKTSLGKPEIADIVASTYRT GPEQLTILPQGLILIRKKNPGGWMGEGLOARKGRQIGMFPANVYKLSFGTSLRTPT EPPSTLAAVACQVITGMVDTYAQNDDELANKQOILNVINKEDPDWKKGVNQQVGLF PSNVYKLTITMDPSEOQMSDILHMLTPERRKQGIHELIVTEENVYNDLQVLEI FOKLMSEELTEKEVAMI PVNMKELTMCIKLLKALVRKMSGEMPKMIGDILS AOLPHMQPIRFGSROLNGALLQOKTDEVDPRPEYKRLANDPRCKGMLSPILKP MORTYRPLIILIKNILENTPENPHDSHLKALAEKABLCSQVNBGVAREKNSDLEMT QAHQVCGLSBQVAFNSVTNCLGPRKRLHSGKLYKASNSNELGFLNEDULLQVTLK PLAGSGDKEVSPKSNLOQMYMKTPIPLNBLVYLPDPGSDBPFIHSIINDRYLTPL AESINERTANQOKIKAASELYIETEEKKREKAYLVRSQATYGLGRLAVNCEGELKP CRSGKSNPYCEVLTMSGSCHTITRTIOTPLNKNNSNOFPIRLDEQVLCITVYERQ FSPDPFGRTIRIVADIKQOQSGKPTKLLLHVEVTEGIYVRDLQLFDEP"				
BASE COUNT	1893 a	1523 c	1639 g	1384 t	
ORIGIN					

Query Match 58.5%; Score 2975; DB 9; Length 6439;
Best Local Similarity 86.0%; Pred. No. 0;
Matches 3367; Conservative 0; Mismatches 515; Indels 33; Gaps 5;

QY 7 GAGAGAGATGAGCGCGCGGAGGCGCGGAGCTTGTGCTCGTAGTACGCGGCT 66
DB 27 GAGAAAGATGAGCGCGCGGAGGAGGAGCTTGTGCTCGTAGTACGCGGCT 86
QY 67 GCGAAGGAGGAGCTCCGAGCGGCTCCGAGAGCGCGGAGGAGGAGCGCGGCT 126
DB 87 CCGAGAGAGAGATCCGAGCGGCTCCGAGAGCGCGGAGGAGGAGCGCGGCT 135
QY 127 GCGAGTGTGTGCGCGGCTCGGAGCTCGGCTTCTGCGCGCGGAGGAGGAGCT 185
DB 136 GCGAGTGTGTGCGCGGCTCGGAGCTCGGCTTCTGCGCGGAGGAGGAGCT 195
QY 186 TTTGTGTGAGGAGCGCGCGCGGAGCGCGGAGGAGGAGGAGGAGGAGGAGG 245
DB 196 TTTGTGTGTGAGGAGCGCGCGCGGAGCGCGGAGGAGGAGGAGGAGGAGG 255
QY 246 AGCTAATAGAACCATGCTCAGTTTCCCACTTTGCTGTGAGTGTGCTGCTG 305
DB 256 AAGTAAACAACCATGCTCAGTTTCCCACTTTGCTGTGAGTGTGCTGCTG 315
QY 306 CATTAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 365
DB 316 CATTAAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 375
QY 366 GGGATTTTAACTGTGATCAAGCGAGAACTTTTTCATCTGAGGTTACTGAGC 425
DB 376 TGGATTTTAACTGTGATCAAGCGAGAACTTTTTCATCTGAGGTTACTGAGC 435
QY 426 TGTCTTAGAGCAAAATGTGGCGCTAGCGGAGCATGAATACATGAGAGATGATCA 485
DB 436 TGTCTTAGAGCAAAATGTGGCGCTAGCGGAGCATGAATACATGAGAGATGATCA 495
QY 486 GGAATTTTTCATAGCCATGAGCTTATCAACTGAAGTACAAAGATCAGTCCCTC 545
DB 496 GGAATTTTTCATAGCCATGAGCTTATCAACTGAAGTACAAAGATCAGTCCCTC 555
QY 546 CACACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 605
DB 556 TCCACTTCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615
QY 606 TATAGAGGAGGATGCTAGCATGCGACCATCAAGCTGTGCTCTGTGCGCAATG 665
DB 616 TATAGAGGAGGATGCTAGCATGCGACCATCAAGCTGTGCTCTGTGCGCAATG 675
QY 666 CATTCAGTGTGTGAAATGTCTCAACCTTAGATCTGTGCTCCCGAGGAGGAGG 725
DB 676 CATTCAGTGTGTGAAATGTCTCAACCTTAGATCTGTGCTCCCGAGGAGGAGG 735
QY 726 TCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 785
DB 736 CCCCCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 795
QY 786 ACCCATGTGCGAAGAGGTTCTTCTTCAAGAGATGTGCTCAAGGCTCAAAAT 845
DB 796 ACCCATGTGCGAAGAGGTTCTTCTTCAAGAGATGTGCTCAAGGCTCAAAAT 855
QY 846 TAAAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 905
DB 856 TAAAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 915
QY 906 TGTGCTCAGTCAATCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 965
DB 916 TGTGCTCAGTCAATCAAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 975
QY 966 GAGTGAACACTTAACAGGTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1025
DB 976 GAGTGAACACTTAACAGGTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1035

QY 1026 GGCCTAGCTGTGCTCAATGAGATCTTTCTGACATTTGATCAAGATGAGGAGG 1085
DB 1036 GGCCTAGCTGTGCTCAATGAGATCTTTCTGACATTTGATCAAGATGAGGAGG 1095
QY 1086 AGAAGATTTTATCTTACGATGACCTTAATGATGTTGATGCTGTGCTGAGC 1145
DB 1096 AGAAGATTTTATCTTACGATGACCTTAATGATGTTGATGCTGTGCTGAGC 1155
QY 1146 GCGCGTCTGCTCCAGAGATTCATCCCTCTTCCAGAGAGGAGGAGGAGGAGG 1205
DB 1156 ACTGTCTGCTCCAGAGATTCATCCCTCTTCCAGAGAGGAGGAGGAGGAGG 1215
QY 1206 GATGTCCGTCAATGATCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCTGCT 1265
DB 1216 TATATCTGTCAATGATCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCTGCT 1275
QY 1266 GATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1322
DB 1276 AGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1335
QY 1323 GAACTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1382
DB 1336 GAACTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1395
QY 1383 GCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1442
DB 1396 GCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1455
QY 1443 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1502
DB 1456 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1515
QY 1503 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1562
DB 1516 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1575
QY 1563 GCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1622
DB 1576 GCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1635
QY 1623 ACTCTGATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1682
DB 1636 ACTCTGATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1695
QY 1683 GACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1742
DB 1696 GACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1755
QY 1743 TCAGATATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1802
DB 1756 TCAGATATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1815
QY 1803 TAGAGAGCTAAGATTTGTGAAATCACCACCTTACAGAGGAGGAGGAGGAGGAGG 1862
DB 1816 TAGAGAGCTAAGATTTGTGAAATCACCACCTTACAGAGGAGGAGGAGGAGGAGG 1875
QY 1863 AATGCTTGAAGAGCTTATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1922
DB 1876 AATGCTTGAAGAGCTTATTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1935
QY 1923 GCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1982
DB 1936 GCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1995
QY 1983 GCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2042
DB 1996 ACTAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2055
QY 2043 GCTGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2102
DB 2056 ACTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2115
QY 2103 ACAGCAACTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2162

Db 2116 GCAACAATCCAGAGCAAAAAGTCCATGAGGCTTAACGATCGAAACGAAAGAACAGA 2175
 Qy 2163 GAGGAAGAGCTCGAGTTAGAGAACAAAAGAGACCTTCAGAGACGATTCAGAAAAG 2222
 Db 2176 ACGAAAGATCATAGATTAGAAAACAAAAGAAAGAACCCAAACAGAGCTCAGGAAAG 2235
 Qy 2223 GGAACAAGATGGCTGAGAGCATGTGCACAGAGAG---GAGCAGCCACGCCCCGGAAC 2279
 Db 2236 GGAACAAGATGGCTGAGAGCATGTGCACAGAGAGAGAGCATCAGAGACCAAGAAAAC 2295
 Qy 2280 CCACGAGAGAGACAGACTGAAGAGGAGAGACAGTGTCAAGAAAGAGAGGAGAGAGAG 2339
 Db 2296 CCACGAAAGAGAAAACGAAAAGGAGAGAGTGTCAAAAAGAAAGATGGCGAGGAAAA 2355
 Qy 2340 AGCCAAAGCCGGAATATGCAAGACAGAGAGTGGCTTTCCATCCGCAATCAGAGCCAGC 2399
 Db 2356 AGGCAAAAGAGAAAGACAAAGACAGCTGGTCCGCTTTCCATCAACCAAGAACAGC 2415
 Qy 2400 TAAAGTGGCCACCCAGGACCCCTGGCTTACCAAGAGAAAGGCCCGCTTACCATTTCCG 2459
 Db 2416 TAAAGCAGCTGTCCAGGACCCCTGGCTTACCAAGAGAAAGGCTCCATTTACATTTCCG 2475
 Qy 2460 ACAGAGAGATGTAAGAGTATATTATACCGAGCGCTTACCCCTTGAATCAGAAATCA 2519
 Db 2476 ACAGAAATATGTAAGAGTGTATTTACCGGCACTGATCCCTTTGAATCAGAAAGCA 2535
 Qy 2520 CGATGATATCACCATCCAGCCAGAGATTAATCTCT-----GTTGATGA 2564
 Db 2536 TGATTAATATCACTATCCAGCCAGAGACATGTCTGTGTAAGAGGAAATGGTGTGAAGA 2595
 Qy 2555 AAGCCAGCTGAGAGCCAGAGATGGCTTGAAGAGAGAGTGAAGAGGAGAACCGAGTGT 2624
 Db 2566 AAGCCAACTGAGAACCCGCTGGCTTGAAGAGAAATTAAGAGAAAGACAGGAGTGT 2655
 Qy 2625 CCCTGCAAACTATGAGAAAAGATTCAGAAAATGAGGTTCCCATCCAGCCAAACAGT 2684
 Db 2656 CCCTGCAAACTATGAGAAAATCCAGAAAATGAGGTTCCCGCTCCAGTGAACAGT 2715
 Qy 2685 GACCATCTGACATCTGCCCCCTGCCCCCAATCTGCTGTGAGACCCCTGCTCTT 2744
 Db 2716 GACTATTCATCATCTGCCCCCTGCCCCCAATCTGCTGTGAGACCCCTGCTCTT 2775
 Qy 2745 GCCAGTGAACCTCTGAGCCCTCCAGCAACCCCAACCAATGGGAGAGCTTCACTTCCAG 2804
 Db 2776 GGCATTAACCTCTTGAAGCCCTCCAGCAACCCCTTAACTGAGCCAGCTTCACTTCCAG 2835
 Qy 2805 GTGGCCAGCAGCTCAACGAGAGCCAGAAACGAGCAACTGGGATACGTGGCCGCTCA 2864
 Db 2836 GTGGCCAGCAGCAGAGATGAGAAACGAGAAACGAGATTAATGGATGGAGAGCCCA 2895
 Qy 2865 GCTTCTCTGACCGTACTGAGTGTGCGGCACTTACCGGAGAGATCAGCTTTATCCCGAC 2924
 Db 2896 GCTTCTCTGACCGTACTGAGTGTGCGGCACTTACCGGAGAGATCAGCTTTATCCCGAC 2955
 Qy 2925 CACAGCCAGCTGCTCTCCCATCTCCGCTCTGAGGAGAGGAGGAGGAGGAGGAGGCT 2984
 Db 2956 CACAGCCAGCTGCTCTCCCATCTCCGCTCTGAGGAGAGGAGGAGGAGGAGGAGGCT 3015
 Qy 2985 ACAAGCCGAGCCCTGTATCCCTGAGAGCCAAAAAGACCACTTAATTTTAAACA 3044
 Db 3016 ACAAGCTCAAGCCCTATATCTTGAAGAGCCAAAAAGACCACTTAATTTTAAACA 3075
 Qy 3045 AAGTACGTCAATCCGCTTGAAGACAGACAGATGTGTGTGTTGAGAAAGTTCAAGG 3104
 Db 3076 AATATATGTCATCACCGTCTGAGAACAGACAGATGTGTGTGTTGAGAAAGTTCAAGG 3135
 Qy 3105 TCAGAAAGGTTGGTCCCAAGCTTGAAGTGAATTAAGGAGGAGGAGGAGGAGGAGGAGG 3164
 Db 3136 TCAGAAAGGTTGGTCCCAAGCTTGAAGTGAATTAAGGAGGAGGAGGAGGAGGAGG 3195
 Qy 3165 CACAAGCATCGATGAGCCCTTACTGAAGAGTCTGTAGTCTTAAGAGAGTGGCTTCCC 3224

Db 3196 TACAAGCATGAAATCTGTTCTTACAGAGTCTCTGATGCTTAAGCGAGTACCTTCC 3255
 Qy 3225 GGGCCGCAAGCCAGCCATTTCCCGAGAGAGTTATTTGCCATGATACATACGAGTTC 3284
 Db 3256 AGCAGCCAGAGCCGCTCTTTGAGAGAAATTTATTTCCATGTACCTTACGAGATTC 3315
 Qy 3285 TGACCAAGAGATTTAACTTTTACAGAGGAGATGTGATTTGTGTTTCAAGAAAGATGG 3344
 Db 3316 TGACCAAGAGATTTAACTTTTACAGAGGAGATGTGATTTGTGTTTCAAGAAAGATGG 3375
 Qy 3345 TGACTGTGACGAGGAAACGTTGGCGCAAGAGTCCGAGCTTCTCCCTTCTAATCTGTGAG 3404
 Db 3376 TGACTGTGACGAGGAAACGTTGGCGCAAGAGTCCGAGCTTCTCCCTTCTAATCTGTGAG 3435
 Qy 3405 GCTTAAGATTCAGAGGAGCTGTGAACTGTGAGGAAACAGAGGAGTTAGAAAAAAC 3464
 Db 3436 GCTTAAGATTCAGAGGAGCTGTGAACTGTGAGGAAACAGAGGAGTTAGAAAAAAC 3495
 Qy 3465 TGAATTTGCCAGGTTATTTGCTTCTTACGCTGTACTGTCTCCGAACTACCTGAGC 3524
 Db 3496 TGAATTTGCCAGGTTATTTGCTTCTTACGCTGTACTGTCTCCGAACTACCTGAGC 3555
 Qy 3525 TCTGCGAGCTGATTTCTGATCCGAAAGAAACCCAGTGTGATGTGGAGAGAGAACT 3584
 Db 3556 CCTGTGAGCTGATTTTGAATCCGAAAAAGAACCCAGTGTGATGTGGAGAGAGAGCT 3615
 Qy 3585 GCAAGCTCGAGGAGAAAGGCGCAGATAGGAGTTTCCAGAAATTAATGTCAAACTTCT 3644
 Db 3616 GCAAGCAGTGGAGAAAGGCGCAGATAGGAGTTTCCAGCTTAATTAATGTAAGCTTCT 3675
 Qy 3645 AAGCCCGGAAACAGCAAAATCAACCCCACTGAGCTTCCAGAGCCAGTGCAGCCAGC 3704
 Db 3676 AAGCCCTGGAGCGAGCAAAATCACTCAACAGAGCCACTTAAGTCAACAGATTAAGCGGC 3725
 Qy 3705 AGTGTCCAGGTGATCCGAGATGATCACTTACACCGCCAGAGATGATCAAGATCAAGCTT 3764
 Db 3736 AGTGTCCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3795
 Qy 3765 CAGCAAGGAGCAGATTCATCAACGCTCCCAACAGAGAGACCCGAGCTGTGGAAAGAGA 3824
 Db 3796 CAAAGAGGCGCAGATTCATCAACGCTCCCAACAGAGAGACCCGAGCTGTGGAAAGAGA 3855
 Qy 3825 AGTCAATGGGCAATTTGGCTTCCCATCAATTAATTAAGTGAACCAAGACATGGA 3884
 Db 3856 AGTCAATGGGCAATTTGGGAGCTTCCCATCAATTAATTAAGTGAACCAAGACATGGA 3915
 Qy 3885 CCCAGCCAGCATG 3899
 Db 3916 CCCAGCCAGCATG 3930

RESULT 8
 AF064244
 LOCUS 7247 bp mRNA linear PRI 21-NOV-1998
 DEFINITION Homo sapiens interseclin long form mRNA, complete cds.
 ACCESSION AF064244
 VERSION AF064244.1 GI:3859854
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 7247)
 Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
 Antonarakis, S.E.
 Two isoforms of a human interseclin (ITSN) protein are produced by
 brain-specific alternative splicing in a stop codon
 Genomics 53 (3), 369-376 (1998)
 TITLE
 JOURNAL MEDLINE
 PUBMED 9799604
 REFERENCE 2 (bases 1 to 7247)
 Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and
 Antonarakis, S.E.

TITLE Direct Submission
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue
Michel-Servet, Geneva 4 CH-1211, Switzerland

FEATURES
source Location/Qualifiers

1. 7247
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.1-q22.2"
/tissue_type="brain"
/dev_stage="fetus"
107. 5272
/product="intersectin long form"
/protein_id="AAC78611.1"
/db_xref="GI:3859855"

CDS

167. 406
/note="encodes EH domain"
misc_feature
767. 1936
/note="encodes EH domain"
misc_feature
2324. 2524
/note="encodes SH3 domain"
misc_feature
2843. 3019
/note="encodes SH3 domain"
misc_feature
3110. 3286
/note="encodes SH3 domain"
misc_feature
3326. 3520
/note="encodes SH3 domain"
misc_feature
3569. 3748
/note="encodes SH3 domain"
misc_feature
3836. 4390
/note="encodes GDS domain"
misc_feature
4649. 4819
/note="encodes PH domain"
misc_feature
4895. 5143
/note="encodes C2 domain"
BASE COUNT 2156 a 1652 c 1747 g 1692 t
ORIGIN

Query Match 57.0%; Score 2898.4; DB 9; Length 7247;
Best Local Similarity 86.4%; Pred. No. 0;
Matches 3244; Conservative 0; Mismatches 491; Indels 21; Gaps 3;

165 GGGGGGTGGGGGTGACGATTTTGTGAGGGGGGGGGGGGGGACCGGGGGGAGT 224
13 GGGGGGTGGGGGTGACGATTTTGTGAGGGGGGGGGGGGGGACCGGGGGGAGT 72
225 AGGCGGTGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 284
73 AGGCGGTGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
285 TGGTACCTGATGATGCTGGGCGCATTAAGTGAAGGAGGAGGAGGAGGAGGAGG 344
133 TGGGAGCTGATGATGCTGGGCGCATTAAGTGAAGGAGGAGGAGGAGGAGGAGG 192
345 CCTTACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 404
193 CCAATGTTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 252
405 CCAATGTTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 464
253 TCAATGTTTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 312
465 CGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 524
313 TGAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 372
525 ACAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 584
373 ACAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
585 TTCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 644
433 TTCAAGTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 492
645 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 704
493 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 552
705 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 764
553 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 612
765 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 824
613 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 672
825 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 884
673 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 732
885 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 944
733 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 792
945 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1004
793 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 852
1005 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1064
853 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
1065 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1124
913 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
1125 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1184
973 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1032
1185 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1244
1033 TGTCTCTGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092

QY	1245	GCCTCAGGAGCCCTGCTCAGAGGATGAGCAGAC--CAGAGAAACTGCTGTGAC	1301
Db	1093	ACCAAGAGGAACCAAGTTTATGAAAGTGAACAAACAATTAGAAAAATTAACCTGTAAAC	1152
QY	1302	ATTTGAAAGTAAAGACGGGAGAACTTCGAGGAGGACAGTGTGAGCTGTGAAACCGCGC	1361
Db	1153	GTTTGAAGTTAATAAAACCGGAGAACTTTGAACTGTGCACTGTGAACTGTGAGAAACGAAG	1212
QY	1362	CCAAACGCTCTTTGAGCAGCAGCGCAAAAGACAGAGCGGTGGCTCAGCTGAGCGCGC	1421
Db	1213	GCAAGCTCTCTTGAAACAGCAGCGCAAGAGCAGGAGCGCCTGCGCCAGCTGAGCGGAGC	1272
QY	1422	CGAGCAGAGAGAGAAAGACGGGACCGCCAGAGACAGAGAGCTGAAACCGCAGCTGAGACT	1481
Db	1273	GGAGCAGAGAGAGAAAGAGCGGTGAACCGCCAGAGAGACCGCAAAAGACAACTGGAACT	1332
QY	1482	GGAGAGACGACTGTGAGAAAGCAGCGGAGCTGGAGCGGCAGCAGAGAGAGAGAGGAA	1541
Db	1333	GGAGAGACAACTGTGAAAAAACCGGAGAGTTAGAACCGCAGAGAGAGAGAGAGAGAA	1392
QY	1542	GGAGATCGAGAGCGCGAGGCGCAAAAACGGGAACTGGAAAGCGCGACAACTTGAATG	1601
Db	1393	AGAAATTTGAGAGGCGAGAGGCGTGCAAAAACGGGAACTTGAAGAGCAACACAACTGAGTG	1452
QY	1602	GGAAACGGAAACCGGAGACAGGAACCTCTGAAATAGAGAAACAAGAGCAGAGGCGACCT	1661
Db	1453	GGAAACGGAAATCCAAAGCAGAACTCTAAATCAAAGAAACAAGAAACAAGAGCAATAGT	1512
QY	1662	GGTCTTGAAGCGAAAGAGGAAGACATGAGAGTTGAGTTAGAAAGCTGGAATGACAAAAA	1721
Db	1513	TGTACTGAAAGCAAAAGAAAAGAACTTTGAAATTTGAATTAAGAGCTCTTAATGATTAATA	1572
QY	1722	GCATCAGCTAGAGAGGAAACCTTCAGGATATCAGGTGTGACTGTGCAACCCAGAGCGCAAG	1781
Db	1573	GCATCAACTAGAGAGGAAACCTTCAGGATATCAGAGTGTGAGTTGACCAACCCAGAGCGCAAG	1632
QY	1782	AATTGAGAGCACGAACAAGTCTAGAGAGCTAAGAAATTCGTAATACCCACTTAACAGCA	1841
Db	1633	AATTGAGAGCACAAACAATCTAGAGAGTGTGAAATTCGCAAAATCACCCACTCTACAGCA	1692
QY	1842	GCAGTTGACAGGAATTCACGCAAAATGCTTGGAAAGCTTAATTCACAGAGAAACAGATCTCAG	1901
Db	1693	ACAATTAACAGGAATCTCAGCAAAATCTTGGAAAGCTTAATTCAGAAAAACAGATCTCAA	1752
QY	1902	TGACCAAGTTAAAAACAAGTCCAGCAGAAACAGTTGCAATAGAGACTGTCTTAAACCCCTCA	1961
Db	1753	TGACCAATTAATAACAAGTTACAGCAGAAACAGTTTGACAGAGATTAACCTGTTTACACTTA	1812
QY	1962	AAGACCTTTGGAACAAAGAGAGCTGCGCGCAGCAGCTCCGGGAGCACTGTGACGAGGT	2021
Db	1813	AAGACCTTTAGAGCAAAAGAACTAGCTCGCAGCAGCACTACGAGCAACTGTGATGAAGT	1872
QY	2022	GGAGAGAGAGACCAAGTCAAAAGCTGCAAGAGATTAATGTTTTCAACAACAGCTGGAAGA	2081
Db	1873	GGAGAGAGAGAACTAATCAATCAAACTACAGAGATTAATTTCAATTAATCAGCTGGAAGA	1932
QY	2082	ACTGAGAGAGATCAATTAGCAAAACAGCAACTCCAGAGAGAGAGTCCCTGAGGCGACGCG	2141
Db	1933	ACTAAGAGAGATACCAATTAAGCAACAATCCAGAGAGCAAAAGTCAATGAGAGCTGAACG	1992
QY	2142	ACTGAGAGAGAAAGACAGAGAGAGAGAGCTGTGAGTTAGAGAGCAAAAGAGAGAGCGC	2201
Db	1993	ACTGAGAGAGAGAAAGAGAGAGAGAGAGATCAATTAAGAAAAACAATAAGAGAGAGC	2052
QY	2202	TCAGAGACAGAGTTCAAGAAAGAGGACAAAGCAATGCTGTGAGACATGTGCACAGAGAG--GA	2258
Db	2053	CCAAAGAGAGAGCTCAGAGAAAGGAGCAACAAGTGTGAGAGATGTGCACAGAGAGAGAGCA	2112
QY	2259	GCAGAGCAGCGCCCGGAAACCCCAAGAGAGGAGCAGACTGAAAGAGAGAGAGAGCAAGTGTGAG	2318
Db	2113	GCATTAAGAGACCAAGAGAAATCTCAAGAGAGAGAGAACTGAAAGAGGAGAGAGAGTGTCA	2172
QY	2319	GAGAGAGAGGCGGAAGAGAGAGCCAGACCGGAATGTGACAGCAAGAGAGTGGCTTTT	2378

Db	2173	AAAGAGATGGCGAGGAAAAAGCCAAACAGAGAGCAACAGACAAAGCTGGGTCCGCTTTT	2232
OY	2379	CCATCCGATCAGAGAGCAGCTAAAGCTGAGCCAGCCAGGCACTCTGCTTACACAGAGAA	2438
Db	2233	CCATCAACACCAAGAACAGAGCTAAGCCAGCTGTCCAGGCACTCGTGTCACTGCAGAAAA	2292
OY	2439	AGCCCCGCTTACCAATTTCTGCACAGAGAGGTGTAAAAAGTGTATATTAACGAGCGCTGT	2498
Db	2293	AGGTCCACTTACATTTCTGCACAGGAAAAATGTAAAAAGTGTATATTAACCGGCACTGT	2352
OY	2499	CCCCTTTGAATCCAGAAAGTCACAGTGCATCATCATCAGCCAGAGATATATGATCAT---	2555
Db	2353	CCCCTTTGAATCCAGAAAGCCATGTATGAATCTACTATCCAGCCAGAGAACATATGATGCT	2412
OY	2556	-----GGTGATGAAACCACAGACTGAGAGCCAGAGATGGCTGTGAGAGAGACT	2603
Db	2413	TAAAGGGGAATGGGTGTGATGAATACCAACTGTAGAGAACCCGGCTGCTGTGAGAGATTT	2472
OY	2604	GAAAAGGAGAGCGGGATGTTCCCTGCACAACTATGACAAAAAGATTCCAGAAAAATGAGT	2663
Db	2473	AAAAGGAAAGCAGAGGTGTTCCCTGCACAACTATGACAGAAAAATCCAGAAAAATGAGT	2532
OY	2664	TCCCATCTCCAGCCAAACAGATACCATCTGACATCTGCCCCCTGCCCCCAACTGGCTCT	2723
Db	2533	TCCCGCTCCAGTGAACCAAGTACATGATTCACAACTGCCCCCTGCCCCCAACTGGCTCT	2592
OY	2724	GGGTGAGACCCCTCTCTCTTTGACAGAGACCTCTTTGAGGCCCTCCAAACCCCTCAACAA	2783
Db	2593	GGGTGAGACCCCGCCCTTTTGGAGATTAACCTCTTCAGAGCCCTCCACAGACCTTAATTA	2652
OY	2784	CTGGGACAGACTTCACTTCCAGTGGCCACAGAGCTCAACGAGAGCCAGAAACGACAA	2843
Db	2653	CTGGGCGAGACTTTCAGCTTCACAGTGGCCCAACAGACCAATATGAGAAACAGAAACGATTA	2712
OY	2844	CTGGGATACGTGGGCGGCTCAGCTTCTCTGACCTGATACCTAGTGTGGCCAGTTACGGCA	2903
Db	2713	CTGGGATGCAATGGGCAAGCCAGCCCTCTCTCAACGTTCCAAAGTGCACGGCCAGTTAAAGCA	2772
OY	2904	GAGATCAAGCCCTTTTACCACAGCAACAGCACTGGTCTCTCCCAACTCTCCGTCCTGGGCA	2963
Db	2773	GAGGTCCGCTTTTACTCAGACCAAGCCACTGGCTCTCTCCCGTCTCTGTGTCTAGGCCA	2832
OY	2964	GGGTGAAAAGGTGAGAGGGCTTACAGAGCCAGACCCCTGTATCCCTGAGAGCCAAAAAGA	3023
Db	2833	GGGTGAAAAGGTGAGAGGGCTTACAGAGCTCAAGCCCTATATCTCTTGAGAGCCAAAAAGA	2892
OY	3024	CAACCACTTAATTTTAACAAAAGTACGTATACCGTTCCTGGAACAGCAACATGTG	3083
Db	2893	CAACCACTTAATTTTAACAAAATGTATCATCACCGTCTGTGAAACAGCAACATGTG	2952
OY	3084	GTGGTTTGGAGAAATTCAGAGGTTCAGAAAGGGTGGTTCCCAAGCTTAACGTAAACATCAT	3143
Db	2953	GTGGTTTGGAGAAATTCAGAGGTTCAGAAAGGGTGGTTCCCAAGCTTAACTGTAAACATCAT	3012
OY	3144	TTCAAGGGCCGTAAAGAAATCCACAAAGCATGTACTGTGCTTACTGAAATCCTGCTAG	3203
Db	3013	TTCAAGGGCCCAATAAGAAAGTCTAACAACTGAAATCTTGAGTTCTTCAAGAGTCTCGCTAG	3072
OY	3204	TCCTAAAGAGTGGCTTCCCGGCGCCAGACCAAGCATTCCTCGAGAGAGATTATTCG	3263
Db	3073	TCCTAAAGCGATGCTCTCCAGACGCAAGACCGGTGTTCCGGGAAAGAAATTAATTCG	3132
OY	3264	CATCTTACATACAGAGGTTCTGAGCAAGAGATTTTAACTTTACAGAAAGGGATGTGAT	3323
Db	3133	CATCTTACATTCAGAGGTTCTGAGCAAGAGATTTTAACTTTACAGAAAGGGATGTGAT	3192
OY	3324	TGTGTTACCAAGAAATGTGATCTGTGACGAGAACGGTGGGCAACAAGTCCGAGT	3383
Db	3193	TTTGTGTTACCAAGAAATGTGATCTGTGACAGAGAACAGTGGGCAACAAGTCCGAGT	3252
OY	3384	CTTCCCTTCTAATATGTAGAGCTTAAAGATTCAGAGGGCTGTGAACTGCTGGAATAAC	3443

```

Db      3253  CTTCCTTCTAAGTATGAGGCTTAAGATTGAGAGGCTCTGGAATGCTGGGAAAC 3312
Qy      3444  AGGAGATTAGAAAAAACTGAAATTCGCCAGATTATGCTTCTAGCGCTGCTAG 3503
Db      3313  AGGAGATTAGAAAAAACTGAAATTCGCCAGATTATGCTTCTAGCGCTGCTAG 3372
Qy      3504  TCCGAGAACATCAGCCCTGCTCTGGGAGCTGATTCGATCCGAAAAAGAACCCAG 3563
Db      3373  CCCGAGAGAGCTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3432
Qy      3564  TGGATGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3623
Db      3433  TGGATGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3492
Qy      3624  AGCAATTAATGCAAACTTCTAAGCCCGGAAACAGCAAAATCAACCCCAATGAGTACC 3683
Db      3493  AGCTAATTAATGCAAACTTCTAAGCCCGGAAACAGCAAAATCAACCCCAATGAGTACC 3552
Qy      3684  CAAAGCCGAGTGCAGCCAGCAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 3743
Db      3553  TAAGTCAAGAGCATGTAGCCGCAAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 3612
Qy      3744  GAAAGTCAAGAGCATGTAGCCGCAAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 3803
Db      3613  GAATGAGCATGAGCTGCTGCTTCAAGAGGCTGATCATCAAGCTCTCAAGAGAGAG 3672
Qy      3804  CCCGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3863
Db      3673  CCTGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3732
Qy      3864  AAAGTCAAGAGCATGTAGCCGCAAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 3899
Db      3733  GAAAGTCAAGAGCATGTAGCCGCAAGTGTGCGAGTGTGCGAGTGTGCGAGTGTGCGAG 3768

RESULT 9
AF032118      4103 bp      mRNA      linear      VRT 13-NOV-1998
LOCUS      Xenopus laevis intersecitin mRNA, complete cds.
DEFINITION
ACCESSION      AF032118
VERSION      AF032118.1 GI:2642624
KEYWORDS
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM      Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Xenopus.
REFERENCE      1 (bases 1 to 4103)
AUTHORS      Yamabhai, M., Hoffman, N.G., Hardison, N.L., McPherson, P.S.,
Castagnoli, L., Cesareni, G. and Kay, B.K.
TITLE      Intersectin, a novel adaptor protein with two Eps15 homology and
five Src homology 3 domains
JOURNAL      J. Biol. Chem. 273 (47), 31401-31407 (1998)
MEDLINE      99030416
PUBMED      9813051
REFERENCE      2 (bases 1 to 4103)
AUTHORS      Hoffman, N.G., Hardison, N.L., Yamabhai, M. and Kay, B.K.
TITLE      Direct Submission
JOURNAL      Submitted (30-OCT-1997) Pharmacology, University of Wisconsin, 1300
University Ave., Madison, WI 53706-1532, USA
FEATURES
source
1..4103
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/cell_type="oocyte"
193..4005
/function="involved in endocytosis"
/note="EH domain and SH3 domain containing protein;
similar to intersectin binding proteins Ibp1, encoded by
GenBank Accession Number AF057285, and Ibp2, encoded by
Rab/Rip, encoded by GenBank Accession Number AF057287"

```

```

/codon_start=1
/product="intersecitin"
/protein_id="AAC73068.1"
/db_xref="GI:2642625"
/translation="MAQGTPEPGNLDIMATITVERAKHDQFHGLKPTAGYITDQA
RNFPLQSLGIPQVLAQIMWALDMNNDGMDQLEFSLAKLILKTLQGLPLSPSM
LKQVAPMAAAVAGGMSGIVGIPPLAAVAPVPMPSIIVGMSPLVSVPLPSPN
GAPAIQSHPAFAHSAATLPGKSGRFAVSGQINIKLQKAGSFVPAPVYEMVPS
SSRLKRLQFNSQDTMGNSLUTGPARTILMOSSLPQSLATINLSLIDQDKITAE
EFTLMHLIDVAMSGQPLPLPPEYIIPSPFRVSGSLSSSVSDQRLPEEPER
EPPQADKLLPVTEFKERENFENRNLLEKRLALLOQREORLEKLEABOERK
REERODERKEQOOLEKORLELERREERREKLEERAKLEEROLEERKE
RROELNORNEOEDIVLKKAKKLEPELELNDKQLEBKLODITCRILTQHEI
ESTNKRRLRIALITLQOOLQESQOILQXMIPEKOSLIDOLKQVQNSLRHSDLLTL
KRALTEKIGROQLBDQDVEYKERTALQETIDVNNQLEKLEIRLNQOFRQKQDFE
TEKIQKELERTSELDLKEDKRLMLEQDLMDRVKQEBERYKFDDEEKREES
VQKCEVKEKPEIQENKPKFHQPEPGGLGQIQPMNTEKAPLITINQDVKVVYRAL
YFPDARSHDELTIEPGDIIMVDESQTPGMGLGELKGTGFPANVABRMSEFSP
TTKPAEFTTAKPTVAPVAPAAFTNTSTNSNMADPSTPTNNNDKVSNDMDT
MAOBSLTVPSAGCHORSAPTPPATVQSSSPVIGQEKVEGLQAOALYPRARKDN
HLNFRKNVITVLEQDDMMWFGVEVGQGMPPKSYVYKLSGPIRKSISTIDTSSEPA
SLKRVSSPAFPALQEBERYISNYTESNEQDLYFOQDLYLVIRKDDDMWTVGEK
TGVPFSNVVRPQDSEAAAGSGKTSGLGKPKBIQVIAVIAVYAAATPQLTLAPOLLIR
KNIPGWMWGELOAGKROQIGFPANVYKLSPTNSTPTPEPKPSTLPTCOVG
MYDYIAONDDELAFSGQVINVLNKEDEDDMKWGLGELNGVGLFSPSVYVLLTMDMSQO
FRLGVKPAQGIPTGDRPFIPLPPRDSGLPNAQAPLPSVMIKRCPFARPCPD
NAVKIKIN"
253..498
/note="encodes Eha domain"
misc_feature
850..1101
/note="encodes EHB domain"
misc_feature
2386..2571
/note="encodes SH3A domain"
misc_feature
2881..3057
/note="encodes SH3B domain"
misc_feature
3148..3324
/note="encodes SH3C domain"
misc_feature
3364..3558
/note="encodes SH3D domain"
misc_feature
3607..3786
/note="encodes SH3E domain"
BASE COUNT      1290 a      920 c      856 t      8 others
ORIGIN
Query Match      40.3%; Score 2047.2; DB 5; Length 4103;
Best Local Similarity 73.6%; Pred. No. 0;
Matches 2713; Conservative 4; Mismatches 932; Indels 39; Gaps 7;
Qy      217  CGAGATGAGGCGTGCATCAGACGAGTGAAGTATAGACCATGAGCTGATCCCA 276
Db      151  CCGTATAGGACCTTACTAGACACTATATGATACACCATGAGCTGATGAACT 210
Qy      277  CTTTGGTGTAGCCTGATGCTGGCCATTAATCTGTGAGAAAGGCCAAGCATGAC 336
Db      211  CCGTTGGGGGTAATTGACATCTGGCCATTAACGTGAGAGAACGAGCTAAACATGAC 270
Qy      337  CAGCGTTCCTTAGCCTGAAGCCGATPAGCGGATTTATTACTGTGATCAAGCGAGAAC 396
Db      271  CAGCGATTCATGGGCTCAAGCAACAGCTGATATATACAGAGGACCAAGCTAGGAAT 330
Qy      397  TTTTTCATCTGGGTTACCTCAGCCTGCTTAGACAAATATGGCGCTAGCGGAC 456
Db      331  TTTTTCCTTCACTCCGCTCCCGAGCTGTGCTGCGACAGATATGGGCTGCTAGC 390
Qy      457  ATGAATAACATGAGAGAGATGATCAATGGAATTTTCATAGCATGAGCTTAA 516
Db      391  ATGAACAATGATGAGAGATGAGCACTGAGGATTTCTCAATAGCATGAATATATCAAG 450
Qy      517  CTGAAGCTACAAAGATATACGCTCCCTCCACACTTCCCTGTCATGAAACAGAACCA 576
Db      451  CTAAACATACAAAGTTACCGCTGCTCTTATTTCCCTTCAATATGTTGAGCAACCG 510
Qy      577  GTGGCTAT-----TTCAAGTGACACAGCATTTGGTATAGAGAGGATTTGATGATGCA 630

```


QY 2791 GACTTCAGTTCCAGTGGCCAGAGCTCAAGAGAGCCAGAAACGACACTGGAGT 2850
 DB 2698 GACTTCAGTTCCAGTGGCCAGAGCTCAAGAGAGCCAGAAACGACACTGGAGT 2757
 QY 2851 ACCTGGGCGGCTCAGCCCTCTCTGACCGTACCTAGTGGCCAGTTACGGCAGAGATCA 2910
 DB 2758 ACATGGGCGGCTCAGCCCTCTCTGACCGTACCTAGTGGCCAGTTACGGCAGAGATCA 2817
 QY 2911 GCCTTACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 2970
 DB 2818 GCTTTTACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 2877
 QY 2971 AAGGTGGAAGGCTTACAGGCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 3030
 DB 2878 AAGGTGGAAGGCTTACAGGCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 2937
 QY 3031 TTAATTTTAAACAAAGTACGTCATCAGCTTCTGGAACAGCAGACATGTGGTGGT 3090
 DB 2938 CTAAATTTTAAACAAAGTACGTCATCAGCTTCTGGAACAGCAGACATGTGGTGGT 2997
 QY 3091 GGAAGAGTTCAAGTTCAGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3150
 DB 2998 GGGGAAGTTCAAGTTCAGAGGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3057
 QY 3151 CCCGTAAAGGAAATCCAGAGCATCGATCTGCGCTACTGAAAGTCTGCTAGTCTAAAG 3210
 DB 3058 CCACTAAAGGAAATCCAGAGCATCGATCTGCGCTACTGAAAGTCTGCTAGTCTAAAG 3117
 QY 3211 AAGAGTCTTCCCGGCGCGGCAAGCCATCCCGGAGAGAGTTTATGTCATGTAC 3270
 DB 3118 AAGAGTCTTCCCGGCGCGGCAAGCCATCCCGGAGAGAGTTTATGTCATGTAC 3177
 QY 3271 ACATAGAGAGTTCTAGAGAGAGAGTTTATGTCATGTCAGAGAGAGTGTGATGTGTT 3330
 DB 3178 ACATAGAGAGTTCTAGAGAGAGAGTTTATGTCATGTCAGAGAGAGTGTGATGTGTT 3237
 QY 3331 ACCAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3390
 DB 3238 ATTAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3297
 QY 3391 TCTAATATGTAAGGCTTAAAGATTCAGAGGCTCTGGAACGTCGTCGGAACAGGAGT 3450
 DB 3298 TCTAATATGTAAGGCTTAAAGATTCAGAGGCTCTGGAACGTCGTCGGAACAGGAGT 3357
 QY 3451 TTRAGAAAAAAGCTGAAATTCGCAAGCTTATGCTTCTGACGTCGTCGTCGTCGTCGTCG 3510
 DB 3358 CTTGGGAAAGAAACCAAGAAATTCGCAAGCTTATGCTTATGCAACCACTGCTCTCGAA 3417
 QY 3511 CAACCTACCGCTGCTGCGGCGAGTTCGATTCGAGGAAAGAAACCAAGTGGATG 3570
 DB 3418 CAGCTAACCTGCTGCTGCGGCGAGTTCGATTCGAGGAAAGAAACCAAGTGGATG 3477
 QY 3571 TGGGAAGAGAACTGCAAGCTCTAGAGGAAAGAGCCAGATAGGCTGTTTCCAGCAAT 3630
 DB 3478 TGGGAAGAGAACTGCAAGCTCTAGAGGAAAGAGCCAGATAGGCTGTTTCCAGCAAT 3537
 QY 3631 TATGTCAAACTTAAAGCCCGGAGCAAGCAAAATCAACCCCACTAGCTAACCAAGCC 3690
 DB 3538 TATGTCAAACTTAAAGCCCGGAGCAAGCAAAATCAACCCCACTAGCTAACCAAGCC 3597
 QY 3691 GCAGTGCAGCCAGCAAGCTGTCAGGAGTTCGAGGAAAGTATGATGATGATGATGATGATGAT 3750
 DB 3598 ACCTCACTTGTCTTCAATGCAAGTTATGCGGATATGATGATGATGATGATGATGATGATGAT 3657
 QY 3751 GACGAATAGCTTTCAGCAAGGCGAGATCATCAAGCTCTCAACAGAGAGAGCCGAGC 3810
 DB 3658 GACGAATAGCTTTCAGCAAGGCGAGATCATCAAGCTCTCAACAGAGAGAGCCGAGC 3717
 QY 3811 TGTGTGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3870
 DB 3718 TGTGTGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3777

QY 3871 ACCACAGCATGAGCCAGCCAGCAAT 3898
 DB 3778 ACCACAGCATGATCTTATGATCAAT 3805

RESULT 10
 BD127168
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

BD127168 2131 bp DNA linear PAT 18-SEP-2002
 Primer for synthesizing full-length cDNA and use thereof.
 BD127168.1 GI:23222113
 JP 2002017375-A/2599.
 Homo sapiens (human)
 Homo sapiens

REFERENCE
 AUTHORS
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.

TITLE
 JOURNAL
 COMMENT

Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2599 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2599
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI,PI
 SHINICHI KOJIMA,
 PI TETSUO OTSUKI,HISASHI KOGA

PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key

FT CDS Location/Qualifiers
 1..2131
 Location/Qualifiers
 1..2131
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 694 a 469 c 526 g 442 t

Query Match 31.7%; Score 1610; DB 6; Length 2131;
 Best Local Similarity 85.5%; Pred. No. 0;
 Matches 1892; Conservative 0; Mismatches 295; Indels 15; Gaps 3;

QY 45 GTTGCTCCGTAGTACGCGGCTCGCAGAGGAGCATCCGAGCGGCTCGGAGCGCGG 104
 DB 1 GTTGCTCCGTAGTACGCGGCTCGCAGAGGAGCATCCGAGCGGCTCGGAGCGCGG 57
 QY 105 GAGGAGAGGAGCGCGGCGCGGAGTGTGTGCGCGGCTGGGACTGGGCTTCTGCC 164
 DB 58 -----CAGAGAGCGCGGCGCGGAGTGTGTGCGCGGCTGGGCTTCTGCC 109
 QY 165 -GCGGCGTGGCGGCTGCACTGATTTGTGAGAGGCGCGCGCGCACCGCGGAGAT 223
 DB 110 AGCGGCGGTAGCGCGGACTGATTTGTCTCGGAGGCGGAGCGGAGCCCGGAGAT 169
 QY 224 GAGGCGGTAGTACGAGAGTGAACCTTAATAGAACCATGCTCAGTTTCCACACTTTG 283
 DB 170 GAGGCGGTAGTACGAGAGTGAACCTTAATAGAACCATGCTCAGTTTCCACACTTTG 229
 QY 284 GTTGAGCTGTGATGTCTGGGCTTAACTCTGAGAGAGAGAGCCAGATGACGAGCT 343
 DB 230 GTTGAGCTGTGATGTCTGGGCTTAACTCTGAGAGAGAGAGCCAGATGATGAGAGT 289
 QY 344 TCCCTAGCCTGAGGCGGATGAGGAGTTTATGATGATGATGATGATGATGATGATGATGAT 403

Db 290 TCACATGTTAAAGCCAAATCTGGATTCAATTAAGTATGATCAAGTAGAACTTTT 349
 Qy 404 TCACATCTGGGTTACCTCAGCTGTCTTACGACAAATATGGCGCTACCGGACATGATA 463
 Db 350 TTCAATCTGGGTTACCTCAGCTGTCTTACGACAAATATGGCGCTACCGGACATGATA 409
 Qy 464 AGGATGAAAGGATGATGATGAAATTTTCATATGCCATGAAAGCTTAACTGAAGC 523
 Db 410 ATGATGAAAGGATGATGATGAAATTTTCATATGCCATGAAAGCTTAACTGAAGC 469
 Qy 524 TACAGGATATCAGTCCCTCCACACTTCCCTGTATGAAACAGCAACAGTGGCTA 583
 Db 470 TACAGGATATCAGTCCCTCCACACTTCCCTGTATGAAACAGCAACAGTGGCTA 529
 Qy 584 TTTCCAGTGCACACGATTTTGTATAGAGAGGATTTCTAGCATGACCACTCAGAGT 643
 Db 530 TTTCCAGTGCACACGATTTTGTATAGAGAGGATTTCTAGCATGACCACTCAGAGT 589
 Qy 644 TTGCTCTGTGCGCAATGGGCTCATTCAGTGTGGAATGTCTCAGCCCTTATATCTT 703
 Db 590 TTGCTCTGTGCGCAATGGGCTCATTCAGTGTGGAATGTCTCAGCCCTTATATCTT 649
 Qy 704 CTGTCTCTGACAGCAGTGCCTCCCTGTGCTAACGCGGCTCTCTCCATACAGCCTC 763
 Db 650 CTGTCTCTGACAGCAGTGCCTCCCTGTGCTAACGCGGCTCTCTCCATACAGCCTC 709
 Qy 764 TTGCTCTGTGCGCATCTCTGACGACATGGCCAAAGATTTCTCTTACAGAGATCTG 823
 Db 710 TTGCTCTGTGCGCATCTCTGACGACATGGCCAAAGATTTCTCTTACAGAGATCTG 769
 Qy 824 GTCCAGGCTCAATTAACCTAAGTTACAGAGGACCAATCATTCAGTGTGCGCAGC 883
 Db 770 GTCCAGGCTCAATTAACCTAAGTTACAGAGGACCAATCATTCAGTGTGCGCAGC 829
 Qy 884 CCCCTCCAGCAGCAGATGGGCTGTGCTCAGTCAATCAAGGCTGAATACAGGCAATT 943
 Db 830 TCCCAACAGTGACAGTGGGCTGTGCTCAGTCAATCAAGGCTGAATACAGGCAATT 889
 Qy 944 TCAACAGCAGCAGCAAACTATGATGACACTTAAACAGTCTCCAGGACAGAACTATTC 1003
 Db 890 TCAATAGTCATGACAAACTATGATGACACTTAAACAGTCTCCAGGACAGAACTATTC 949
 Qy 1004 TCATCAATCAAGTTAACCCCGGCTCAGCTGGCTTCAATATGAAATCTTTCTGACATTG 1063
 Db 950 TTATCAATCAAGTTAACCCCGGCTCAGCTGGCTTCAATATGAAATCTTTCTGACATTG 1009
 Qy 1064 ATCAAGATGAAAACTCACTGCAAGAAATTTATCTAGCTATGCACTTAATTGATG 1123
 Db 1010 ATCAAGATGAAAACTTACAGCAAGAAATTTATCTAGCTATGCACTTAATTGATG 1069
 Qy 1124 CCATGCTGTGTAGCGCATGCGCGCGCTGCTGCTCCAGAAATACATCCCTCTTCTTCA 1183
 Db 1070 CCATGCTGTGTAGCGCGCATGCGCGCGCTGCTGCTCCAGAAATACATCCCTCTTCTTCA 1129
 Qy 1184 GAAAGATTCGCTCCGCGAGTGGATGTCCGATATAGCTCTTCTGATGATGAGG 1243
 Db 1130 GAAAGATTCGATTCGCGAGTGGATGTCTGATATAGCTCAACATCTGATGATGAGG 1189
 Qy 1244 TGCCTGAGAGCGCTGCTGCAAGATGAGCAGCAGC--CAGAGAAAGAACTGCTGTA 1300
 Db 1190 TACCAAGAGAACAGTTTAAAGAAATGAACAACAATTAAGAAAGAAATTAACCTGTAA 1249
 Qy 1301 CATTTGAAGATTAAGAAAGCGGGAATTTGAGCGAGGCGAGTGTGAGCTGGAAGACGCC 1360
 Db 1250 CATTGGAAGATTAAGAAAGCGGGAATTTGAACTGTGCAACTGTGAACGTGGAAGACGAA 1309
 Qy 1361 GCCAAGCGCTTTGGAGCAGCAGCGGAAGAGCAGAGAGGTTGGCTCAGCTGAGCGCG 1420
 Db 1310 GCCAAGCGCTTCTTGAACAGCAGCGGAAGAGCAGAGAGGCTTGGCCAGCTGAGCGCG 1369
 Qy 1421 CCGAGCAGAGAGAGAAAGACCGGAGCGCCAGAGAGAGAGGCGCAAGCGCAGCTGAGC 1480
 Db 1370 CCGAGCAGAGAGAGAGAGCGTGAAGCGCCAGAGAGAGAGAGCGCAAGAAAGCAACTGGAAC 1429

Qy 1481 TGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAGCGGCGACGAGAGAGAGAGAGAG 1540
 Db 1430 TGGAGAGCAGCTGGAGAGCAGCGGAGCTGGAGCGGCGAGAGAGAGAGAGAGAG 1489
 Qy 1541 AGGAGATCAGAGAGCGGAGCGGCAAAACGGGAATCGAAAGCGCAGCAACTTGAAT 1600
 Db 1490 AAGAAATTTAGAGAGCGGAGCGTGCAGAAACGGGAATTTGAAGGCAACCACTTGA 1549
 Qy 1601 GGGAGCGGAACCGGAGCAGAGAACTCTGAATCAGAGAAACAGAGACAGAGAGCGC 1660
 Db 1550 GGGAGCGGAATCGAAGCGCAAGAACTTAATCAAAAGAAAGAAAGAGAGAGCTAG 1609
 Qy 1661 TGTCTGGAAGGCAAGAGAGAGAGCTGGAGTTTGAATGAGAGCTCTGATAGCAAAA 1720
 Db 1610 TTGTACTGAAGAGCAAGAGAGAGAGCTTGAATTTGAATTTGAAGCTCTTAATGAT 1669
 Qy 1721 AGCATCAGCTAGAGAGAGAGAGAGAGATTCAGAGTGTGCACTGGCAACCCAGAGC 1780
 Db 1670 AGCATCAACTAGAGAGAGAGAGAGAGATTCAGATGTGATGATGACCAACCAAGC 1729
 Qy 1781 AAATTGAGAGCAG 1840
 Db 1730 AAATTGAGAGCAG 1789
 Qy 1841 AGCAGTTGAG 1900
 Db 1790 AACATTAAG 1849
 Qy 1901 GTGACAGATTAAACAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960
 Db 1850 ATGACCAATTAAACAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1909
 Qy 1961 AAAAGCCTTGAAG 2020
 Db 1910 AAAAGCCTTGAAG 1969
 Qy 2021 TGGAG 2080
 Db 1970 TGGAG 2029
 Qy 2081 AACTGAGAGAGATCATAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2140
 Db 2030 AACTGAGAGAGATCATAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2089
 Qy 2141 GACTGAG 2182
 Db 2090 GACTGAG 2131

RESULT 11
 AK074554
 LOCUS 2131 bp mRNA linear PRI 03-SEP-2002
 DEFINITION Homo sapiens cDNA FLJ90073 f18, clone HEMBA1004110, highly similar to Homo sapiens interseotin short form mRNA.
 AK074554
 ACCESSION AK074554.1 GI:22760070
 VERSION 1
 KEYWORDS Oligo capping, f18 (full insert sequence).
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 Isegai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y., Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Ninomiya,K.
 TITLE NED0 human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2131)
 AUTHORS Isegai,T. and Otsuki,T.

TITLE
JOURNAL

Direct Submission
Submitted (25-MAR)

2002) Takao Isogai, Helix Research Institute,

COMMENT

FEATURES

Source

1. 2131

FROM 2005

```
/usr/bin/perl -e "print 'Perl\n'"
```

15

BASE COUNT
ORIGIN

694 a 469 c 526 g 442 t

Query Match
Best Local

31.7%;	Score 1610;	DB 9;	Length 2131;
85.5%;	Pred: No. 0;		

QY	45	GTTCCTCCGTAAGTACGGGAGGCTCCGAGAGGACATCCCGAACGGGCTTCGGGAGCGGCGG	104
Db	1	GTTCTCCGTATGTAACGGCGGCTCCGGAGGAAGATCCCGAACGGGCTTCGGGAGCGGA	57
QY	105	GAGGACAGCAGCGCGGCGGGGATGTGTGTGCGCGGCTCCGGACTGGCGTTCTCGC	164
Db	58	-----CAAGAGCGCGGCGGGGATGTGTGTGCGGGGCTCCGGCTCTGTGGTCCCTCC	109
QY	165	-GCGGCGGCGGGCGTGCACTGATTGTGTGAGAGGGCGGCGCGGCGACCCGCGAGAGAT	223
Db	110	AGCGGCGCGTACGCGGACCTGATTTGTCTCGGGCGGCGACGCGGACCCGCGGAGAT	169
QY	224	GAGGCGTCGATCAGCAAGGTGAACGTATATAGAACCATGGCTCAGTTTCCACACTTTTCG	283
Db	170	GAGGCGTCGATTTACCAAGTAAAGTAAACAGAACCATGGCTCAGTTTCCACACTTTTCG	229
QY	284	GTGTATGCTGGGATGTCTGGGCGCATATCTGTGAGAGAAAGGCGCAAGATGACACAGCT	343
Db	230	GTGGCAGCTGGGAATCTGGGCGCATATCTGTAGAGAAAGCGAAGCATATCAGCAGT	289
QY	344	TCCTTAGCTGAAACCCGATAGCGGATTTATTACTGGTGATCAAGCGAGAACTTTT	403
Db	290	TCCATATGTTTAAACCAATATCTGTGATTCATTCGTGATTCMAGCTATGAAACTTTT	349
QY	404	TCCAATCTGGGTAACTCAGCCTGTCTTATGACAAATATGGGCGCTAGCGGACATGATA	463
Db	350	TTCAATCTGGGTAACTCMAACCTGTTTATGACAGATATGGGCACTAGCTACATGATA	409
QY	464	ACGATGGAAGATGGATCAAGTGGAAATTTCCATAGCATAAGCTTATCAACTGAAGC	523
Db	410	ATGATGGAAGATGGATCAAGTGGAGTTTCCATAGCTATATAAATTATCAACTGAAGC	469
QY	524	TACAAGATATCAGTCCCTCCACACTTCCCTGTTCATGAAAACAGCAACAGTGGCTA	583
Db	470	TACAAGATATCAGTACCTCTGCACTTCCCTGTATGAAAACAGCAACATTTGCTA	529
QY	584	TTTTCAGTGACCCAGCATTTTGTATAGAGGGAATTTCTAGCATGCACTCACAAGTGTG	643
Db	530	TTTTCAGTGACCCAGCATTTTGTATAGAGGATATGCAAGATGCAACCGCTTACAAGTGTG	589
QY	644	TTGCTCTCGTACCAGTGGCTTCATTTCAAGTTGTGGAATGCTCCACCTCTAGATATTT	703
Db	590	TTGCTCTCAGTCCCAATGGGATTCATTTCAAGTTGTGGAATGCTCCCAACCTCTAGATATTT	649
QY	704	CTGTCTCTCCAGACAGAGTGCCTCCCTGGCTAAAGGGGCTTCCCGTCAATACAGCTC	763

Db	650	CTGTTC	CCCA	CAGAC	GTGTG	CCCCC	CTGG	CTAA	CGGGG	CTCCCC	CTGT	TAT	CA	CC	TC	709
Qy	764	TG	CT	GTG	CG	TTT	CG	CA	CT	CTG	CAG	CC	CA	AT	GG	823
Db	710	TG	CT	GT	CA	TTT	GT	CA	CT	CTG	CAG	CC	CA	AT	GG	769
Qy	824	GT	CC	AG	GT	CA	CA	TT	TA	CA	CT	TA	AG	TT	TA	883
Db	770	GT	CC	AG	GT	CA	CA	CT	TA	CA	CT	TA	AG	TT	TA	829
Qy	884	CC	CT	CC	AG	CG	CA	GA	AT	GG	CG	CT	GT	CA	TA	943
Db	830	T	CC	CA	CG	AT	GG	CG	CA	GA	AT	GG	CG	CT	GT	889
Qy	944	T	CA	CA	CG	CC	CA	GA	CA	AA	CT	TA	GA	GT	GA	1003
Db	890	T	CA	TA	GT	CA	TA	GA	CA	AA	CT	TA	GA	GT	GA	949
Qy	1004	T	CA	TA	GT	CA	TA	GA	TT	TA	CA	CC	CA	GG	CT	1063
Db	950	T	TA	TA	GT	CA	GT	CA	AA	CT	TA	GA	TT	TA	CA	1009
Qy	1064	A	T	CA	GA	AT	GG	AA	CT	CA	TA	GA	AA	TT	TA	1122
Db	1010	A	T	CA	GA	AT	GG	AA	CT	TA	GA	AA	TT	TA	CA	1065
Qy	1124	C	A	T	G	T	C	G	T	G	A	C	C	A	T	1183
Db	1070	C	T	A	G	T	C	G	G	C	A	C	A	C	T	1129
Qy	1184	G	A	A	G	A	T	T	G	A	T	T	G	A	T	1243
Db	1130	G	A	A	G	A	T	T	G	A	T	T	G	A	T	1189
Qy	1244	T	G	C	T	G	A	G	A	G	C	G	T	C	T	1300
Db	1190	T	A	C	A	G	A	G	A	C	A	G	T	T	A	1249
Qy	1301	C	A	T	T	G	A	A	G	A	T	T	G	A	A	1360
Db	1250	C	G	T	T	G	A	A	G	A	T	T	G	A	A	1308
Qy	1361	G	C	A	A	G	C	T	T	G	A	G	C	A	G	1420
Db	1310	G	C	A	A	G	C	T	T	G	A	G	C	A	G	1369
Qy	1421	C	C	G	A	G	C	A	G	A	T	T	G	A	A	1480
Db	1370	C	C	G	A	G	C	A	G	A	T	T	G	A	A	1429
Qy	1481	T	G	G	A	G	C	A	G	A	T	T	G	A	A	1540
Db	1430	T	G	G	A	G	C	A	G	A	T	T	G	A	A	1488
Qy	1541	A	G	A	G	A	T	T	G	A	A	T	T	G	A	1600
Db	1490	A	G	A	A	T	T	G	A	A	T	T	G	A	A	1549
Qy	1601	G	G	G	A	C	G	A	A	T	T	G	A	A	T	1660
Db	1550	G	G	G	A	C	G	A	A	T	T	G	A	A	T	1609
Qy	1661	T	G	G	C	T	G	A	A	G	A	T	T	G	A	1720
Db	1610	T	T	G	A	C	T	G	A	A	T	T	G	A	A	1669
Qy	1721	A	G	C	A	T	C	A	G	A	T	T	G	A	A	1780
Db	1670	A	G	C	A	T	C	A	G	A	T	T	G	A	A	1729
Qy	1781	A	A	A	T	T	G	A	G	A	T	T	G	A	A	1840

Db 1730 AATTGAGAGCAAAACAAATCTAGAGATTGAGATTGCCAAATCAACCATCTACAGC 1789
Qy 1841 AGCAGTTGACAGAAATCTTACGAAATGCTTGGAAAGACTTATTCAGAGAAACAGTACTCA 1900
Db 1790 AACCAATTACAGAAATCTCAGCAAAATGCTTGGAAAGCTTATTCAGAAAAAGATACTCA 1849
Qy 1901 GTGACAGTAAACAACTCCAGCAGAAACAGTTCATAGAGACTCGCTTCTTACCTCA 1960
Db 1850 ATGACCAATTAAACAACTTCAGCAGAAACAGTTCATAGAGACTTCATGACTTCA 1909
Qy 1961 AAAGAGCTTGAAGCAAAAGAGCTGCGCCGAGCAGACTCCGAGAGAGCTGAGCAGG 2020
Db 1910 AAAGAGCTTGAAGCAAAAGAGCTGCGCCGAGCAGACTTCAGAGAGCACTGATGAG 1969
Qy 2021 TGGAGAGAGAGACCGCTCAAGCTGCGAGAGATTGATGATTTTCAACACAGCTGAGG 2080
Db 1970 TGGAGAGAGAGACCGCTCAAGCTGCGAGAGATTGATGATTTTCAACATATGAGCTGAGG 2029
Qy 2081 AACTGAGAGATACATATGACAAACAGCAACTCCAGAAACAGAGTCCCTGAGGAGCGC 2140
Db 2030 AACTGAGAGATACATATGACAAACAGCAACTCCAGAAACAGAGTCCATGAGGCTGAGC 2089
Qy 2141 GACTGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTGAAGTTAG 2182
Db 2090 GACTGAGAGAGAGAGAGAGAGAGAGAGAGATCATGATGATG 2131
RESULT 12
LOCUS BD158570 2131 bp DNA linear PAT 17-JAN-2003
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD158570
VERSION BD158570.1 GI:27864328
KEYWORDS JP 2002191363-A/13413.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2131) Ota,T., Isega,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.
Primer for synthesizing full-length cDNA and use thereof
J Patent: JP 2002191363-A 13413 09-JUL-2002;
JOURNAL HELIX RESEARCH INSTITUTE
COMMENT OS Homo sapiens (human)
PN JP 2002191363-A/13413
PD 09-JUL-2002
PI 28-JUL-2000 JP 2000280990
PI TOSHIO OTA,TAKAO ISOgai,TETSUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TETSUJI OTSUKI
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10.
PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
FEATURES
source FT CDS Location/Qualifiers
1..2131
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 691 a 448 c 572 g 420 t
ORIGIN
Query Match 27.6%; Score 1404.2; DB 6; Length 2131;
Best Local Similarity 81.0%; Pred.No.1.2e-311;
Matches 1740; Conservative 0; Mismatches 283; Indels 126; Gaps 4;
7 GAGGAGAGATGAGCGCGCGGAGGAGCGCGAGCTTGCTCCGATGATGCGCGCT 66
|||||

Db 105 GAGAGAGTGAAGCGCGCAGGAGGAGGAGCTGATGTTGCTCCGTACTGACGCGCT 164
Qy 67 CGCAGGAGACATCCGAGGCGGCTCCGAGAGCGCGGAGGAGGAGGAGGCGGCG 126
Db 165 CGGAGAGAAATCCGAGGCGGCTCCGAGAGCG-----GACAGAGAGCGGCG 213
Qy 127 GGGATGTTGCGCGCGCTGCGGACTGCGCTGCTCGC-GCGGCGTCCGCGGCTGACTGA 185
Db 214 GGGATGTTGCGCGGCTGCGGCTGCGCTGCTCGCTCCCTCCAGCGGCGGCTGAGCGGACTGA 273
Qy 186 TTTGTGAGAGGCGCGCGCGCAGCCGCGGAGATGAGGCTGATGAGCAAGTTGA 245
Db 274 TTTGTCCCTGCGGCGGCGCGGCGGCGGCGGCGGAGATGAGGCTGATGAGCAAGTTGA 333
Qy 246 ACGTAATGAGAACATGGCTCAGTTTCCACACCTTCGGGATGAGCTGATGCTGGG 305
Db 334 AAGTAACAGAACATGGCTCAGTTTCCACACCTTCGGGATGAGCTGATGCTGGG 393
Qy 306 CATTAAGTGAAGAGAGAGGCGCAAGCATGACAGCAAGTTCTTAAAGCCGATGAC 365
Db 394 CATTAAGTGAAGAGAGAGAGGCGCAAGCATGACAGCAAGTTCTTAAAGCCGATGAC 453
Qy 366 GGGATTTATTAATCTGATGATGAGGAGAACTTTTTCATCTGGTTACTGAGCC 425
Db 454 TGATTCATTAATCTGATGATGAGGAGAACTTTTTCATCTGGTTACTGAGCC 513
Qy 426 TGCTTGAACAAATATGAGGCGCTAGCGGACATGAATATGAGAGATGATGAT 485
Db 514 TGTTTAGCAGAGATATGAGGCGCTAGCGGACATGAATATGAGAGATGATGAT 573
Qy 486 GGAATTTTCATAGCCATGAGAGCTTATCAACTGAACTCAAGATATGAGCTCCCTC 545
Db 574 GGAATTTTCATAGCCATGAGAGCTTATCAACTGAACTCAAGATATGAGCTCCCTC 633
Qy 546 CACACTTCCCTCTCATGAAACAGCAACAGTGGCTATTTCCAGTCCACCATGTTGG 605
Db 634 TGCACTTCCCTCTCATGAAACAGCAACAGTGGCTATTTCCAGTCCACCATGTTGG 691
Qy 606 TATGAGAGGATTCATGAGTGCACACATGCAAGCTGTTGCTCTGCAATGGGCTC 665
Db 692 ----- 691
Qy 666 CATTCAGTTGTTGAGATGCTCCACCCTTATCTTCTGCTCCACAGCAGTGC 725
Db 692 -----CGAGTGTGCC 702
Qy 726 TCCCTGCTAAGCGGCGCTCCCTCCGTCATACAGCTGCTGCTGCTGCTGCTGCTG 785
Db 703 CCCCCTGCTAAGCGGCGCTCCCTCCGTCATACAGCTGCTGCTGCTGCTGCTGCTG 762
Qy 786 AGCCACATGCGCAAGAGTCTCTCCATGAGATCTGCTGAGGCTCAGCAATTAACAC 845
Db 763 AGCCACATGCGCAAGAGTCTCTCCATGAGATCTGCTGAGGCTCAGCAATTAACAC 822
Qy 846 TAAATTAAGAGGAGCAATCATTCATGATGCGCAGCGCCCTCCAGCAGAGATGGGC 905
Db 823 TAAATTAAGAGGAGCAATCATTCATGATGCGCAGCGCCCTCCAGCAGAGATGGGC 882
Qy 906 TGTGCTGATGATCAAGAGCTGAAATACAGGCAATTAACAGCAACGCAAAACTAT 965
Db 883 TGTGCTGATGATCAAGAGCTGAAATACAGGCAATTAACAGCAACGCAAAACTAT 942
Qy 966 GAGTGAACATTAACAGGCTCCCGAGGAGAACTATCTCATGCAATCAAGTTAACCCA 1025
Db 943 GAGTGAACATTAACAGGCTCCCGAGGAGAACTATCTCATGCAATCAAGTTAACCCA 1002
Qy 1026 GGTCAAGTGGCTCAATATGAAATCTTTCGATGATGATGATGATGATGATGATGAT 1085
Db 1003 GGTCAAGTGGCTCAATATGAAATCTTTCGATGATGATGATGATGATGATGATGAT 1062
Qy 1086 AAGAAATTTATCTTACATGACATTAATGATGATGATGATGATGATGATGATGAT 1145
Db 1063 AAGAAATTTATCTTACATGACATTAATGATGATGATGATGATGATGATGATGAT 1122

QY	1146	GCCGCTCTGGCTCCAGAAATACCCCTCTTCCTCAAGAGTGGCTCCGCGACATGG	1205
Dp	1123	ACCTGCTCTGCTCCAGAAATACCTTCCTTTTGAAGAGTTCATCTGGCAATGG	1182
QY	1206	GATGTCCTCATTAAGCTCTTCTTGTGTGATCAGAGGCTGCTGAGAGCCGTGTCAG	1265
Dp	1183	TATATCTGTCAATTAAGCTCAACATCTGTATGATCAAGAGCTTACCAAGAGAACGATTTT	1242
QY	1266	GGATGAGAGCAAGC---CAGAGAGAACTGCTGTGTGACATTTGATTAAGAACGGGA	1322
Dp	1243	AGATGAACAAACAAATTGAAAAAGAAATTACCTGTAACTGTTTGAAGATTAAGAACGGGA	1302
QY	1323	GAATCTTCAGGAGGACAGTGTGTGAGCTGAGAAAGGCGCCGCAAGGCGCTCTTGGAGCA	1382
Dp	1303	GAATCTTTGAACGTGGCAACTGTGAACTGTGAAGAAACGAAGGCAAGCTCTCTCTGAAACAGCA	1362
QY	1383	GCGCAAAAGCAGAGAGCGGTTGGCTCAGCTGAGAGCGCCGAGCAGAGAGAGAAAGACG	1442
Dp	1363	GCGCAAGAGCAGAGAGCGCTTGCCCAAGCTGAGAGCGGCGAGCAGAGAGAAAGAGCG	1422
QY	1443	GGAGCGCCAGAGAGCAGAGGCGCAAGCGGCACCTGAGCTGTGAAGAAAGCAGCTGTGAAGCA	1502
Dp	1423	TGAGCGCCAGAGAGCAAGAGCGCAAAAGCAACTGTGAATGTGAAGAACCACTGTGAAGCA	1482
QY	1503	GCGGAGCTTGAGCGGCGCAGAGAGAGAGAGAGAGAGAGATTCGAGAGCGCGAGGC	1562
Dp	1483	GCGGAGCTTAAACAGCAGAGAGAGAGAGAGAGAGAGAGAAATTGAGAGCGCAGAGGC	1542
QY	1563	CGCAAAACGGAACTGTGAAGGCGAGCGCAACTTGATGTGGACCGAAACCGAGACAGGA	1622
Dp	1543	TGCAAAACCGGAACTTTGAAGAGCAACGCAACTTGAGTGGGAGCGGAATCGAAGCAAGA	1602
QY	1623	ACTCTGTAATAGAGGAACAAGAGCAGAGAGGCAACGTGTTCTTGAAAGGCAAGAGGA	1682
Dp	1603	ACTACTAATCAAAAGAAACAAAGAACAAAGAGACATATTTGTACTGAAAGCAAAAGAAA	1662
QY	1683	GACTCTGAGTTTGATGTTAGAAAGCTCTGAATGACAAAAAGCATCAGCTAGAGAGAAAACT	1742
Dp	1663	GACTTTGGAATTTGAATTTAGAAAGCTCTTAATATGATTAAGAAAGCATCACTAGAGGAAACT	1722
QY	1743	TCAGAGATTCAGGTGTGCACTGGCAACCCAGAGCGCAAAATTGAGACACGAACAAGTC	1802
Dp	1723	TCAAGATATCAGATGTGCATTTGACCACCCCAAGGCAAGAAATTTGAGAGCACAAAATC	1782
QY	1803	TAGAGAGCTAGAAATTTGCTGAATCAACCACTTACAGAGAGATTCAGGAATCTCAGCA	1862
Dp	1783	TAGAGATTTGAAATTTGCCGAAATCCCACTCTACAGCACAATTTACAGGAATCTCAGCA	1842
QY	1863	AATGCTTGAAGACTTATTTCCAGAGAAACAGATCTCAGTACCAAGTTTAAACAAGTCCA	1922
Dp	1843	AATGCTTGAAGACTTATTTCCAGAAAAACAGATCTCAATGACCAATTTAAACAAGTTCA	1902
QY	1923	GCAGAACAGTTTGCATAGAGACTGCTTTTAACTCTCAAAAAGCCTTTGGAAGCAAAAGGA	1982
Dp	1903	GCAGAACAGTTTGCACAGAGATTTCACTTTTACCTTTAAAAAGCCTTTAGAGCAAAAGA	1962
QY	1983	GCTGGCCCGGAGCAGCTCCGAGAGCAGCTGAGAGAGGTGAGAGAGACAGAGTCAAA	2042
Dp	1963	ACTAGCTTCGAGCACTTACAGAGACCAACTGATGAAAGTGAAGAAAGAACTTAAGTCAA	2022
QY	2043	GCTGAGAGATTTGATGTTTTCACAAACACAGCTGAAGGAACGTGAGAGATATACAA	2102
Dp	2023	ACTACAGAGATTTGATGATTTTCAATTAATCACTGAAGGACTTAAGAGAAATATACAAATTA	2082
QY	2103	ACAGCAATCTCAGAAAGCAGAGSTCCCTGGAAGCGGAGCTTGAAGCAG	2151
Dp	2083	GCAACAACTCCAGAAAGCAAAAGTCCATGGAAGCTGAAAGACTGTAAACAG	2131

RESULT 13			
AK027846			
LOCUS			
AK027846	2131 bp	mRNA	linear
			FRI 01-AUG-2002

DEFINITION	Homo sapiens cDNA FL11940 fis, clone PLACE1010942, highly similar to Homo sapiens interseclin long isoform (ITSN) mRNA.
ACCESSION	AK027846
VERSION	AK027846.1 GI:14042823
KEYWORDS	oligo capping; fis (full insert sequence).
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 Isogai,T., Ohtani,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Ieshashi,T., Fujimori,K., Tanai,H., Kimeta,M., Matanabe,M., Hirooka,S., Ishii,S., Kawai,Y., Saito,K., Yamamoto,U., Wakamatsu,A., Nakamura,Y., Negahari,K., Masuno,Y. and Kanehori,K.
TITLE	NEDO human cDNA sequencing project
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 2131)
AUTHORS	Isogai,T. and Otsuki,T.
TITLE	Direct Submission
JOURNAL	Submitted (10-MAY-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisaerazu, Chiba 292-0812, Japan (E-mail:genomics@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES	Location/Qualifiers
source	1..2131
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="PLACE1010942"
	/cissue_type="Placenta"
	/clone_11b="PLACE1"
	/note="Cloning vector: PME18SFL3"
BASE COUNT	691 a 448 c 572 g 420 t
ORIGIN	
Query Match	27.6%; Score 1404.2; DB 9; Length 2131;
Best Local Similarity	81.0%; Pred. No. 1.2e-311;
Matches 1740; Conservative	0; Mismatches 283; Indels 126; Gaps 4
QY	7 GAGGAGGATGAGAGCGCGCGGAGGCGCGCCAGCTTGTTGCTCCGTAGTACGGCGCT 66
Db	105 GAGAAAGTGAAGGCGCGCAGGAGGAGGAGCGCTAGCTTGTTGCTCCGTATGAGGCGCT 164
QY	67 CGCAAGGAGGATCCCGAGCGGCGCTCCGGAGCGGCGGAGGCGAGCGAGCGGCGGCGG 126
Db	165 CGGAGGAAAGATCCCGAGCGGCGCTCCGGAGCG-----GACAGAGAGCGGCGG 213
QY	127 GGGATGTTGCGCGGCTGCGGACTCGGCGTTCTCTGC-CGCGCGTGGCGGCTGCACTGA 185
Db	214 GGGATGTTGCGGCGGCTGCGGCTCTGCGTCCCTCCAGCGGCGGTGAGCGGCACTGA 273
QY	186 TTTGTGTAGAGGGCGCGCGCGCGCACTCGGCGTTCTCTGC-CGCGCGTGGCGGCTGCACTGA 245
Db	274 TTTGTCTCCGTGGGCGCGAGCGCGGACCGCGCCGGAGATGAGGCGTCAATTAGCAAGTTA 333
QY	246 AGCTAATAGAACATGAGCTCAAGTTCCACACCTTTTGGGTAGCTGAGCTGATGTTCTGGG 305
Db	334 AGCTAATAGAACATGAGCTCAAGTTCCACACCTTTTGGGTAGCTGAGCTGATGTTCTGGG 393
QY	306 CATTACTGTGAGAGAGGCGCAAGCATGACAGCAAGTTCCTTAGCCTGAAGCCGATAGC 365
Db	394 CATTACTGTGAGAGAGGCGCAAGCATGACAGCAAGTTCCTTAGCCTGAAGCCGATATC 453
QY	366 GGGATTATTATCTGTGTCAAGCGAGGAATTTTTTTTCAATCTGGTTACTCAGCC 425
Db	454 TGGATTATTAATCTGTGTCAAGCTAGAGAACTTTTTTTTCAATCTGGTTACTCAGCC 513

```

426 TGTCTTACGACAAATATGGGCGCTAGCGGACATGAATACGATGGAGAGTGAATCAAGT 485
427 TGTCTTACGACAAATATGGGCGCTAGCGGACATGAATACGATGGAGAGTGAATCAAGT 486
514 TGTCTTACGACAAATATGGGCGCTAGCGGACATGAATACGATGGAGAGTGAATCAAGT 573.
486 GGAATTTTCATAGCATGAGTCAATGATCAAACTGAAGTACAGAGATATACGCTCCCTC 545
574 GGAATTTTCATAGCATGAGTCAATGATCAAACTGAAGTACAGAGATATACGCTCCCTC 633
546 CACACTTCCCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACCAAGCATTTGG 605
634 TGCACCTCCCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACCAAGCATTT-- 691
606 TATAGAGGAGATGTAGATGTCACCACTCAAGCTGTGCTCTGTGCCAATGGGCTC 665
692 ----- 691
666 CATTCAGTGTGGATGTCTCAACCTTAGTATCTTCTGTCCCTCAGCAGCAGTGC 725
692 -----GCAAGCTGTGC 702
726 TCCCTGCTAAAGGGGCTCTCCCTGTCATGAGCTCTGCTGCTTGGCATCTGC 785
703 CCCCCTGGCTAAAGGGGCTCTCCCTGTCATGAGCTCTGCTGCTGCTGCTGCTGCTGC 762
786 AGCCCATGGGCAAGAGTCTCTGTCAGAGATCGGTCCAGGGTCAATTAACAC 845
763 AGCCCATGGGCAAGAGTCTCTGTCAGAGATCGGTCCAGGGTCAATTAACAC 822
846 TAAAGTTACGAGAGGACATCAATGATGTCGCGCAGCGCCCTCCAGAGCAGAAATGGGC 905
823 TAAATTTACGAGAGGACATCAATGATGTCGCGCAGCGCCCTCCAGAGCAGAAATGGGC 882
906 TGTGCTCTGATCATGAGCTGAAATACAGGCAATTAATCAAGCAGCAGCAAACTAT 965
883 TGTGCTCTGATCATGAGCTGAAATACAGGCAATTAATCAAGCAGCAGCAAACTAT 942
966 GAGTGGACATTAAGGCTCCAGGCAAGAACTATTCATGATGATGATGATGATGATGATGAT 1025
943 GAGTGGACATTAAGGCTCCAGGCAAGAACTATTCATGATGATGATGATGATGATGATGAT 1002
1026 GAGTGGACATTAAGGCTCCAGGCAAGAACTATTCATGATGATGATGATGATGATGATGAT 1085
1003 GAGTGGACATTAAGGCTCCAGGCAAGAACTATTCATGATGATGATGATGATGATGATGAT 1062
1086 AGAAGATTTATCTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1145
1063 AGAAGATTTATCTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1122
1146 GCGCGTCCGCTCCAGATATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1205
1123 ACCTGCTGCTCCAGATATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1182
1206 GATGCTCCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1265
1183 TATATCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1242
1266 GATGCTCCGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1322
1243 AGATGAACAAACATTAAGAAAGAAATTAACCTGTAACGTTGAAGATTAAGAAACGGGA 1302
1323 GAACCTTCCGAGGAGAGTGTGAGAGTGGAGAAAGGCGGCAAGGCTCTGGAGACAGA 1382
1303 GAACCTTCCGAGGAGAGTGTGAGAGTGGAGAAAGGCGGCAAGGCTCTGGAGACAGA 1362
1383 GCGCAAGAGCAGAGCGGCTGCTGCTGAGTGGAGCGCGCGCAGAGCAGAGAGAGAGAGCG 1442
1363 GCGCAAGAGCAGAGCGGCTGCTGCTGAGTGGAGCGCGCGCGCAGAGCAGAGAGAGAGAGCG 1422
1443 GAGAGCGCAGAGCAGAGCGGCTGCTGCTGAGTGGAGCGCGCGCGCAGAGCAGAGAGAGAGCG 1502
1423 TGAGCGCGCAGAGCAGAGCGGCTGCTGCTGAGTGGAGCGCGCGCGCAGAGCAGAGAGAGAGCG 1482

```

```

QY 1503 GCGGAGCTGAGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562
DB 1483 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542
QY 1563 CGCAAAACGGGAACTGGAAAAGCAGCAGCAACTTTGAAATGGGAAACCGGAGACAGGA 1622
DB 1543 TGCAAAACGGGAACTGGAAAAGCAGCAGCAACTTTGAAATGGGAAACCGGAGAGAGGA 1602
QY 1623 ACTGCTGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1682
DB 1603 ACTGCTGAATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662
QY 1683 GACTCTGAGATTTGAGTTAGAGCTGTAATGACAAAAGAGATCAGTAGAGAGAGAACT 1742
DB 1663 GACTCTGAGATTTGAGTTAGAGCTGTAATGAGATTAATGAGATTAATGAGAGAGAACT 1722
QY 1743 TCAGATATCAGAGTGTGCTGCTGAGCACTCCAGAGCAGAGAAATTTAGAGCAGCAAGTC 1802
DB 1723 TCAGATATCAGAGTGTGCTGAGTGAACCAACCAAGCAGAGAAATTTAGAGCAGCAAGTC 1782
QY 1803 TAGAGCTGAAGATTTGCTGAATCAGCACTTAACAGAGAGTGGAGAGATTCAGCA 1862
DB 1783 TAGAGCTGAAGATTTGCTGAATCAGCACTTAACAGAGAGTGGAGAGATTCAGCA 1842
QY 1863 AATGCTTGAAGACTTATTCAGAGAAACAGATACAGTACAGTGAACAGTTAAACAGTCCA 1922
DB 1843 AATGCTTGAAGACTTATTCAGAGAAACAGATACAGTACAGTGAACAGTTAAACAGTCCA 1902
QY 1923 GCAGAACAGTTTGCATAGAGATCTGCTTTTAACTTCAAAAGAGCTTTGAGAGAGAGGA 1982
DB 1903 GCAGAACAGTTTGCATAGAGATCTGCTTTTAACTTCAAAAGAGCTTTGAGAGAGAGGA 1962
QY 1983 GCTGCGCGGAGAGAGCTCCGCGGAGCAGCTGAGAGAGTGGAGAGAGAGAGAGAGAGAG 2042
DB 1963 ACTAGCTGCGAGAGAGCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
QY 2043 GCTGCGAGAGATTTGATTTTCAACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2102
DB 2023 ACTAGAGAGAGATTTGATTTTCAATTAATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
QY 2103 ACAGCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2151
DB 2083 GCAACCAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

RESULT 14
HSU61166
LOCUS HSU61166 3241 bp mRNA linear PRI 23-JUN-1996
DEFINITION Human SH3 domain-containing protein SH3P17 mRNA, complete cds.
ACCESSION U61166
VERSION U61166.1 GI:1438932
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Sparkes,A.B., Hoffman,N.G., McConnell,S.J., Fowlkes,D.M. and Kay,B.K.
TITLE Cloning of ligand targets: systematic isolation of SH3 domain-containing proteins
JOURNAL Nat. Biotechnol. 14 (6), 741-744 (1996)
MEDLINE 96294438
PUBMED 9630982
REFERENCE
AUTHORS Pirozzi,G., McConnell,S.J., Uveges,A. and Fowlkes,D.M.
TITLE Direct Submision
JOURNAL Submitted (18-JUN-1996) CYTOGEN Corp., 307 College Road East, Princeton, NJ 08540, USA
FEATURES
source
1..3241
/organism="Homo sapiens"
/mol_type="mRNA"

```



```

/db_xref="taxon:9606"
/tissue_type="bone marrow"
37.1599
CDS

```

```

/db_xref="taxon:9606"
/tissue_type="bone marrow"
37..159
/codon_start=1
/product="SH3 domain-containing protein SH3P17"
/protein_id="PAC50592.1"
/db_xref="GI:1438933"

```

[illegible]

BASE COUNT	994 a	756 c	702 g	789 t
ORIGIN				

Query Match	24.6%;	Score 1252.6;	DB 9;	Length 3241;
Best Local Similarity	69.8%;	Pred. No. 9.2e-277;		
Matches 2132; Conservative	0;	Mismatches 604;	Indels 317;	Gaps 20;

QY	2112	CCAGAGACAGAGCTCCCTGGAGCGCCGATGAGGACGAAAGGCGAGAGGAAAG	217
Db	21	CCAGAGCAAAAGTCATGAGAGGCTGAACGACTGAACAGAAAGAACAAAGAAAT	80
QY	2172	CCTGAGTTTAAGAAAGCAAAAGAGAGCGCTCAGAGACGAGTCCAGAAAGGCAACACA	223
Db	81	CATGAAATTAGAAAACAAAAGAAAGAGCCCAAGACGAGTCCAGAAAGGAGCAACCA	140
QY	2232	ATGCGTCGAGCATGTGCAGCAGAG--GAGCAGCCAGCCCCCGAAACCCACGAGGA	2288
Db	141	GTGCGTGAGCATGTGCAGCAGAGAGAGAGCATCAGAGACCAAGAAAACCTCCAGAGA	200
QY	2289	GGACAGACTGAAGGAGAGACAGTGTACGAAAGAGAGGCGAGAGAGAGACCAACC	2348
Db	201	GGAAAAACTGAAAAGGAGAGAGTGTAAAAAGAAAGATGACGAGAAAAAGGCAACA	260
QY	2349	GGAATGCAAGACAAAGCAGAGTCGGCTTTTCCATCCGATCAGGAGCCAGCTTACCTGAC	2408
Db	261	GGAGGCAACAAACAAAGCTGGGTCCGCTTTTCATCAACAACAAAGACCAAGTACCGAC	320
QY	2409	CACCAAGGCACTCTGCTTCAACAAGAAAGGCCGCTTACCATTTCTGACAGAGAG	2468
Db	321	TGTCAGGCACTCTGCTTCACTGCAAGAAAAAGTGTCACTTACCATTTCTGACAGAGAAA	380
QY	2469	TGTAAAGTGTATATTAACGAGGCGTACCCCTTTGAATCCGAATCCAGATGAGAT	2528
Db	381	TGTAAAGTGTGTATTAACCGGCGACGTACCCCTTTGAATCCGAATCCAGATGAGAT	440
QY	2529	CACCATCCAGCAGAGAGATATATGTCATGTTGATGAATGAACCAAGCTGAGAGCCAGATG	2588
Db	441	CACATATCCAGCAAGAGACATATGTCATGTTGATGAATGAACCAAGCTGAGAGACCCTG	500
QY	2589	GCTTGGAGAGAGCTGAAGAGGAGAACCGGATGTTCCCTGCAAACTATGCAAAAGAT	2648
Db	501	GCTTGGAGAGAAATTAAGAGAGAACCGGATGTTCCCTGCAAACTATGCAAGAAAT	560
QY	2649	TCAGAAAAATAGAGTCCCACTCCAGCCAAACCAAGTGCACGATCGACCCCTGC	2708
Db	561	CCAGAAAAATAGAGTCCCGCTCCAGTGAACCAAGTGCATGATTCAGATCTGCCCTGC	620
QY	2709	CCCCAACTGGCTCTGCGTGAAGCCCTGCTCTTTCAGTGAACCTTCTGAGCCCTC	2768
Db	621	CCCCAACTGGCTCTGCGTGAAGCCCCGCCCCCTTTGGACATTAACCTTCAAGAGCCCTC	680
QY	2769	CACAACCCCAACAACCTGGGCAAGCTTCAAGTTCACAGGAGGCCAGACACTCAACAGAA	2828
Db	681	CACAGACCCCTTAATACTGGGCGAGCTTACGCTCAAGTGGCCCAACACACAGATGAAGA	740
QY	2829	GCCAGAAACGACACTGGAGTACGTGGCGGCTCAGCTTCTCTGACCGTACTAGTGC	2888

Db	741	ACGAGAAACGGATTA	CTGGAGTGCATGGGCA	GGCCGACCCCTCTTCA	CCCTTCCAAATGC	800
OY	2889	TGGCAGATTACGG	CAGAGATCAGCCTTTAC	CCCGACACAGCACTGG	CTCTCCCATC	2348
Db	801	CGGCAGATTAAAG	CAGAGGTCCGCTTTAT	CTCCAGCCACGGCACT	GGCTCTCCGCCCTC	860
OY	2949	TCCCGTCTTGSGG	CAGGGGTAAAGGTGA	AGGGCTTCAAGCGAC	GCCTGTATCCCTG	3008
Db	861	TCCGTGTCTAGGG	CCAGGGGTAAAGGTGA	GGGGCTTCAAGCTCA	AGCCCTATATCCTTG	920
OY	3009	GAGAGCCAAAAA	ACACCACTTAAATTTT	TAAAGTGAAGTCA	TACCGTTCTGGA	3068
Db	921	GAGAGCCAAAAA	ACACCACTTAAATTTT	TAAAGTGAAGTCA	TACCGTTCTGGA	980
OY	3069	ACACCAAGACATG	TGTGTGTTTGGAGAG	TTTCAAGGTCAAGAG	GGTGTGTTCCCAAGTC	3128
Db	981	ACACCAAGACATG	TGTGTGTTTGGAGAG	TTTCAAGGTCAAGAG	GGTGTGTTCCCAAGTC	1040
OY	3129	TTACGTAAAACT	CAATTTTCAAGGGCC	CGTAAAGAAATCC	CAAGCATGATCTG	3188
Db	1041	TTACGTAAAACT	CAATTTTCAAGGGCC	CGTAAAGAAATCC	CAAGCATGATCTG	1100
OY	3189	TGAAAGTCTGTG	TATGTCTTAAAGAG	TGTGCTTCCCGGCG	CCCAAGCACCATTCC	3248
Db	1101	AGAAAGTCTGTG	TATGTCTTAAAGAG	TGTGCTTCCCGGCG	CCCAAGCACCATTCC	1160
OY	3249	AGAAAGTATTAT	TGCGCATATACAT	CTACGAGAGTTC	TGACACAAAGAGATTT	3308
Db	1161	AGAA-----				1164
OY	3309	GCAAGGGATGTG	ATGTGTGTTTCA	CAAGAAAGATGTG	ATGTA	3368
Db	1165	-----				1164
OY	3369	CGACAAGTCCG	AGACTTCCCTTCT	TAACTATGTGAG	CTTAAAGATTCA	3428
Db	1165	-----				1164
OY	3429	AACGTGTGGG	AAAAACAGAGAG	TTTAGSAAAAAAC	CTGAAATTTGCC	3488
Db	1165	-----				1164
OY	3489	CTACGCTGCTA	TGAGTCCGGAACA	CACTCACTGAG	CTGCGGACAGTAT	3548
Db	1188	ATACACGCGCA	CCGGCCCCGAGAG	CTCACTCTCGCC	CTGTCACTGATTTG	1247
OY	3549	GAAAAAGAAC	CCAGGTGATGTGTGG	GAAGAGAACTG	CAAGCTCGAGG	3608
Db	1248	AAAAAAGAAC	CCAGGTGATGTGTGG	GAAGAGAACTG	CAAGCTCGAGG	1307
OY	3609	GATGTGGTGTG	TTTCCAGCAAAAT	TATGTCAAACTT	TAAAGCCCGG	3668
Db	1308	GATGTGGTGTG	TTTCCAGCAAAAT	TATGTCAAACTT	TAAAGCCCGG	1367
OY	3669	CCCAACTGAG	CTACCCAAAGCC	AGAGGACGACAG	AGTGTGCAGGTG	3728
Db	1368	TCCAAACGAG	CCACTTAAGTCAA	CAGCATTTAG	CGGCAAGTGTCC	1427
OY	3729	CGATTACAC	CGCCAGACGATG	ACGAATTAG	CTTCAAGCAAA	3788
Db	1428	CGATTACAC	CGCCAGAGAAATG	ACGATGAGCT	GTGCTTCAACAA	1487
OY	3789	CTTCAACAA	AGAGAGACCCG	CACTGTGTGAA	AGAAAGTCA	3848
Db	1488	CTTCAACAA	AGAGAGACCCG	CACTGTGTGAA	AGAAAGTCA	1547
OY	3849	CCCATCCAA	TTATGTAAAGCTG	ACACAGATG	AGACCCAGCACA	3908
Db	1548	CCCATCCAA	TTATGTAGTGA	AGCTGACACAG	ACATGACCCAG	1607
OY	3909	TGTGCAT	TCCCGCTCAG	CTTGAAGTCTT	CAAGAGACCACTAT	3968

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:17:39 ; Search time 23 Seconds
(without alignments)
2231.438 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPFGSLDVAIVE.....VGLFPSNYKLTMDPSQ 1213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*\n2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*\n3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*\n4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*\n5: /cgn2_6/prodata/2/1aa/6CTUS_COMB.pep:*\n6: /cgn2_6/prodata/2/1aa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2058.5	32.8	462	4 US-08-630-915A-38	Sequence 38, Appl
2	1388.5	22.1	509	4 US-08-630-915A-194	Sequence 194, App
3	816.5	13.0	248	4 US-08-630-915A-40	Sequence 40, Appl
4	486.5	7.8	896	1 US-08-095-737-2	Sequence 2, Appl
5	486.5	7.8	896	1 US-08-480-145-2	Sequence 2, Appl
6	486.5	7.8	896	2 US-08-477-389-2	Sequence 2, Appl
7	486	7.8	897	1 US-08-095-737-4	Sequence 4, Appl
8	486	7.8	897	1 US-08-480-145-4	Sequence 4, Appl
9	486	7.8	897	2 US-08-477-389-4	Sequence 4, Appl
10	382	6.1	1162	2 US-08-728-323A-2	Sequence 2, Appl
11	382	6.1	1162	4 US-09-298-568-2	Sequence 2, Appl
12	382	6.1	1162	4 US-09-410-399-2	Sequence 2, Appl
13	381	6.1	1698	4 US-08-056-200-94	Sequence 94, Appl
14	381	6.1	1898	2 US-08-800-644-94	Sequence 94, Appl
15	333	5.3	553	1 US-08-475-894-2	Sequence 2, Appl
16	333	5.3	553	1 US-08-484-710-2	Sequence 2, Appl
17	333	5.3	553	2 US-08-484-709-2	Sequence 2, Appl
18	333	5.3	553	3 US-08-474-697-2	Sequence 2, Appl
19	325	5.2	659	4 US-08-671-354-2	Sequence 2, Appl
20	322	5.1	59	4 US-08-630-915A-133	Sequence 133, App
21	314	5.0	62	4 US-08-630-915A-135	Sequence 135, App
22	309	4.9	56	4 US-08-630-915A-134	Sequence 134, App
23	303.5	4.8	611	4 US-09-216-393B-81	Sequence 81, App
24	301	4.8	1231	4 US-08-714-741-41	Sequence 41, Appl
25	298.5	4.8	864	4 US-08-714-741-40	Sequence 40, Appl
26	297	4.7	2101	1 US-08-466-390-4	Sequence 4, Appl
27	297	4.7	2101	1 US-08-470-950-4	Sequence 4, Appl

28	297	4.7	2101	1 US-08-467-781-4	Sequence 4, Appl
29	297	4.7	2101	2 US-08-483-924-4	Sequence 4, Appl
30	297	4.7	2101	3 US-09-452-294-1	Sequence 1, Appl
31	296	4.7	57	4 US-08-630-915A-136	Sequence 136, App
32	296	4.7	2101	1 US-08-195-487-4	Sequence 4, Appl
33	296	4.7	2101	5 PCT-US93-06160-4	Sequence 4, Appl
34	283	4.5	1805	1 US-07-853-913-2	Sequence 2, Appl
35	281	4.5	683	6 5210183-3	Patent No. 5210183
36	281	4.5	788	4 US-08-630-915A-30	Sequence 30, Appl
37	280.5	4.5	1375	4 US-09-722-139-2	Sequence 2, Appl
38	280.5	4.5	1375	4 US-09-721-832-2	Sequence 2, Appl
39	280.5	4.5	1375	4 US-09-721-689-2	Sequence 2, Appl
40	275	4.4	1360	3 US-09-393-569-2	Sequence 2, Appl
41	275	4.4	1360	4 US-09-579-664B-14	Sequence 14, Appl
42	275	4.4	1360	4 US-09-645-456A-34	Sequence 34, Appl
43	275	4.4	1360	4 US-09-425-324A-34	Sequence 34, Appl
44	275	4.4	1360	4 US-09-645-791-34	Sequence 34, Appl
45	274.5	4.4	1354	3 US-08-685-871-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-630-915A-38
; Sequence 38, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-630-915A-38
Query Match 32.8%; Score 2058.5; DB 4; Length 462;
Best Local Similarity 79.3%; Pred. No. 4.8e-138;
Matches 391; Conservative 13; Mismatches 16; Indels 71; Gaps 1;

```

QY 721 PMSTTEKPLTISAQSVKVVYVYRALYFPESSSHDEITIQODIYMWDESGTGERGWI 780
DB 41 FGLPQKKPLTISAQSVKVVYVYRALYFPESSSHDEITIQODIYMWDESGTGERGWI 100
QY 781 ELKGTGFPANVAEKIPENEPVTPAKPVDTLTSAPAKLALRETPALPTTSSEPTTP 840
DB 101 ELKGTGFPANVAEKIPENEPVTPAKPVDTLTSAPAKLALRETPALPTTSSEPTTP 160
QY 841 NNMADFSSTWPSSEKPEPTDNDTMAAOPSITVPSAQOLRORSAFTPATATGSSPSVL 900
DB 161 NNMADFSSTWPSSEKPEPTDNDTMAAOPSITVPSAQOLRORSAFTPATATGSSPSVL 220
QY 901 GQGEVVEGLQALYPMRACKONHNFNKSDVITVLBOODMMWGEVQOGMPKSYVX 960
DB 221 GQGEVVEGLQALYPMRACKONHNFNKSDVITVLBOODMMWGEVQOGMPKSYVX 280
QY 961 LISGVRKSTSIDTPTSPASLAKRVASPAKPAIPGEFTAMYTESSBOGDLTFQOQD 1020
DB 281 LISGVRKSTSIDTPTSPASLAKRVASPAKPAIPGEFTAMYTESSBOGDLTFQOQD 318
QY 1021 VIVVTKKQDMWGTGVGDKSGVFPSPNYVRLKDSBGSSTAGTGLGKKPELIAQVIASVAA 1080
DB 319 -----EIAQVIASVAA 329
QY 1081 TGPBOLTLAPQOLILIRKKNPGGMEGELQARGKROIGMPANVYKLSGTSKITPTE 1140
DB 330 TGPBOLTLAPQOLILIRKKNPGGMEGELQARGKROIGMPANVYKLSGTSKITPTE 389
QY 1141 LPTAVQAVCVIGMYDTAQNDELAFSGQIINVINKEDPDMMKGEVSGQVGLFPSN 1200
DB 390 LPTAVQAVCVIGMYDTAQNDELAFSGQIINVINKEDPDMMKGEVSGQVGLFPSN 449
QY 1201 YVKLTMDPSQO 1213
DB 450 YVKLTMDPSQO 462

```

RESULT 2

```

US-08-630-915A-194
; Sequence 194, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 194:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 509 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-630-915A-194

```

```

Query Match 22.1%; Score 1388.5; DB 4; Length 509;
Best Local Similarity 48.9%; Pred. No. 1.9e-90;
Matches 287; Conservative 85; Mismatches 124; Indels 91; Gaps 14;

```

```

QY 635 ERKSELEKQK---EDARRVQERDKWLEHVOEESOPRRKHEEDRLKREDSVRKE 690
DB 6 ERKSELEKQKLEDEARAKAKQCKENIMKENIRKEE-----EKQRLQOEKTEQIKQ 59
QY 691 AERAKPEMDQKQRLPHQEPKALQOAPWSTTEKPLTISAQSVKVVYVYRALYFPE 750
DB 60 AERAKPEMDQKQ-----RETASV-----LVNYRALYFPE 86
QY 751 SRSHDEITIQODIYMWDESGTGERGWIAGELKGTGFPANVAEKIP--ENEVPTPAKP 808
DB 87 ARNDEMSFNSGDIITQDEKTVGEPGMLYSGFQGNFGMFCNVEKPKPSSENE----- 139
QY 809 VTDLTSAPAPKALRETPALPTVTSSEPTTPNNMADFSSTWPSSEKPEPTDNDTMAA 868
DB 140 -----KAVSPKAL-----LPTVLSLATS-----TSSEPLSNQOPASVTDYON-VS 180
QY 869 QPSLTVPSAGLQORSAFTPATATGSSPSPTLGGGEVVEGLQALYPMRACKONHNFN 928
DB 181 FSNLTVNTSWQ--KSAFRTVSPG--SVSPHGGQVVENLKALQCSWAKKUNHNFN 237
QY 929 KSDVITVLEQODMMWGEVQOGMPKSYVYKLSGVRKSTSIDTPTSPASLAKRVAS 988
DB 238 KHDITVLEQODMMWGEVQOGMPKSYVYKLSGVRKSTSIDTPTSPASLAKRVAS 289
QY 989 --PAKPAIPGEFTAMYTESSBOGDLTFQOQDVIVTKKQDMWGTGVGDKSGVFPSPN 1046
DB 290 KKPISAAYSVEEYIALYPSVVEPGDLTFEGEILVLTQDGBMWTGSGDRGIFPSN 349
QY 1047 YVRLKDSBGSSTAGTGLGKKPELIAQVIASVATGEOLTLAPQOLILIRKKNPGGME 1106
DB 350 YVRLKDSBGSSTAGTGLGKKPELIAQVIASVATGEOLTLAPQOLILIRKKNPGGME 409
QY 1107 GELQARGKQOIGMPANVYKLSGTSKITPTELPTAVQAVCVIGMYDTAQNDEL 1166
DB 410 GELQARGKQOIGMPANVYKLSGTSKITPTELPTAVQAVCVIGMYDTAQNDEL 462
QY 1167 LAFSGQIINVINKEDPDMMKGEVSGQVGLFPSNYVYKLTMDPSQO 1213
DB 463 LAFSGQIINVINKEDPDMMKGEVSGQVGLFPSNYVYKLTMDPSQO 509

```

RESULT 3

```

US-08-630-915A-40
; Sequence 40, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP

```

```

/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/630,915A
/ FILING DATE: 03-APR-1996
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mirock, S. Leslie
/ REGISTRATION NUMBER: 18,872
/ REFERENCE/DOCKET NUMBER: 1101-174
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 790-9090
/ TELEFAX: (212) 869-8864/9741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 40:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 248 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ US-08-630-915A-40

```

```

Query Match          13.0% Score 816.5; DB 4; Length 248;
Best Local Similarity 58.4%; Pred. No. 2.6e-50;
Matches 153; Conservative 38; Mismatches 54; Indels 17; Gaps 3;

QY 954 PFKSVYKLISGVRKSTSIDTPTSPASLKRVAS--PAKPAIPGEPIFYATYESSHQ 1011
DB 2 FFAAGTKIIPGSEVK-----REPEALVYAAVANKPISAAYSGEPIYALYPISSVVP 53

QY 1012 GDLTQOGDIVYVTKKDGWMTGTVDKSGVPPSNVRLKDSGSGTAGKTSGLKKPEI 1071
DB 54 GDLTFEGEILVTKDGEWMTGSIIDRSIGIPPSNVKPKDQESFSGASKSGASNNKPEI 113

QY 1072 AGVIVSYATGPEOITLARGOLILRKRPAGMWEGELQARGKKOIGFPPANVYKLSL 1131
DB 114 AGVTAIVYASGEOSLAPGOLILIKNTSGMWGELQARGKKRGKGFPAHVYKLSL 173

QY 1132 GTSKITPELPTAVQPAVCQVIGMYDTAQNDDDELAFSGQIINVLNKEDPMMWKGYS 1191
DB 174 SERRATPAHP-----VCQVIAMVDYANNDELSPKGLINVMNKDPTDMWQGEIN 226

QY 1192 GQVGLFPSSNYVYKLTIDMPSQQ 1213
DB 227 GVTGLFPSSNYVMTTIDSPSQ 248

```

```

RESULT 4
US-08-095-737-2
/ Sequence 2, Application US/08095737
/ Patent No. 5487979
/ GENERAL INFORMATION:
/ APPLICANT: DiFiore, Pier P
/ TITLE OF INVENTION: A Substrate for the Epidermal Growth
/ TITLE OF INVENTION: Factor Receptor Kinase
/ NUMBER OF SEQUENCES: 4
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Knobbe, Martens, Olson & Bear
/ STREET: 620 Newport Center Drive, Sixteenth Floor
/ CITY: Newport Beach
/ STATE: California
/ COUNTRY: United States of America
/ ZIP: 92660

```

```

/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/095,737
/ FILING DATE: 19930722
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Israelson, Ned A
/ REGISTRATION NUMBER: 29,655
/ REFERENCE/DOCKET NUMBER: NI060,001A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 235-8550
/ TELEFAX: (619) 235-0176
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 896 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-095-737-2

```

```

Query Match          7.8% Score 486.5; DB 1; Length 896;
Best Local Similarity 20.5%; Pred. No. 4.3e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

QY 15 WAITVERAKHDQQLSLKPIAGFITDQARNFFQSGLPQVLAQIWMALADMMNDGRMD 74
DB 122 WAKVEDRAKYDAIYDLSLSPVNGFLSGDKVKPKVLLNSKLPVILGRVWELSDIDDMGD 181

QY 75 QVESIANKLTKLQGIQQLSTLPVWKQDVAISSAPAFGIGIASNPPLTAVAPVM 134
DB 182 RDEFAVMFLVYCALE-----XEPVPM 203

QY 135 GSIPVGMSPVLVSSVPPAAPPPLANGAPVYIQLPFAFAPAAVTPKSSFSRSGSQL 194
DB 204 -----SLPPLVVP-----SKR 215

QY 195 NTKLQKASFDVASAPPAEWAPOSRLKYROLFNHDKTWSGHLTGPOARTIMQSL 254
DB 216 KT-----WVSPAEKAKYDEIFLTKDKMDGVSGLGEVVEIFLTKGL 257

QY 255 POAQLASTWNISDIDQDKLTAEFTLMLHLDVAM-SGQPLPVLPPEYIPPSRRVRS 313
DB 258 PSTLAAHTWSLCDTDCCKLSKQCPALAFHLISQKLKIGIDPPHYLTPEMTPPS--DRA 314

QY 314 GSGMVISSSSVYDQRLPEPSSSEDEQPEKKLPVTFEDKKRNFERSGVELEKRRQALIE 373
DB 315 SLQKNITGSSPV-----ADPSAIK-----ELDTLNNEIVD 344

QY 374 QQRKEOERLAQLERAEOERKEREROEBAKQLLELEKOLEKORELEEROEERREIERR 433
DB 345 LQREK-----NNVEQDLKEKE-----DTIKQRTSEVQ--DLQDEVQRE 380

QY 434 EAAKELEBOROLEMERNRQELLNQRNKEDEGTVALKARRKTLFELBALNDKKHQLLEG 493
DB 381 NTNLOKLOAQKO-----QVQELIDE-----LDEQAQALE 410

QY 494 KLQDIRCLATROEISTNKSRELRITAEITLQOQLOESQOMGLRIPKQILSDQKO 553
DB 411 QKEVRRKCAEBAQLISLK-----ABLTQSBSQISTYEBELAK-----AREELSR 456

QY 554 VQONSILRDSILTLTALAEKELAQOLREQLDEVERETRS-----KLQEDIVFNNOJK 607
DB 457 LQQETAE-----LEESVESGAQQLFPLQNHQDSQGISSSQWMLMEXKDLNNSQLN 510

QY 608 ELREIHS-----KQOLQKRELEAARLKQKQERKSLTL--- 641
DB 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVEQSNSLESPIHQESPASSPILPSG 570

QY 642 EKQKEDAQRRVQERKQYLEHVQCEQRPKPKHEDELDKEDSVRKKEABER----- 694

```

Db 571 VTDENEVTTATYKVCSELDN-----NRHSKEEDPFNVDSSTLGPVADTULDFPQS 622
 Qy 695 -----AKPEMODKOSRLFHPHOEPK-----LATOAPMSTTEKGLTISAQESV 738
 Db 623 DPFVGSDFPKDDPFCKIDPFQGDPPKSDPPASDCFFRQSTDPFATSTSDPFAANSSI 682
 Qy 739 KVVYRRALYPFESRSHDEITIQPGDIYVWDESQT-----GEPMGLG----- 780
 Db 683 TSV-----ETLKNDPFAFGTVAASDASIDPPASVFGNSFGGFPADFSTLSKVN 734
 Qy 781 ---ELKGTGMFPAN-----YAEKIPENEVPTPAKPVTDLTSAAPKALRETAPLP 830
 Db 735 NEDPRSATSSSVSNVITKAVFETSVKSEDEPALP-----PKIGTPRCPPLP 785
 Qy 831 -----VTSSEP-----STTPNNMADPSSSTWSSSNEKPEFTDMDT 865
 Db 786 PGKSINKLDSPPDKLNDPFPQPPGNDSPKXOPFEMFCDFTSATTTNKADPSPNFAN 845
 Qy 866 WAAOPS 871
 Db 846 FSAIYS 851

RESULT 5 US-08-480-145-2

; Sequence 2, Application US/08480145
 ; Patent No. 5717067
 ; GENERAL INFORMATION:
 ; APPLICANT: Difiore, Pier P
 ; APPLICANT: Fazioli, Francesca
 ; TITLE OF INVENTION: A Substrate for the Epidermal Growth
 ; FACTOR RECEPTOR KINASE
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: United States of America
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/480,145
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 530
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/095,737
 ; FILING DATE: 22-JUL-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Israelien, Ned A
 ; REGISTRATION NUMBER: 29,655
 ; REFERENCE/DOCKET NUMBER: NIH060.001A
 ; TELEPHONE: (619) 235-8550
 ; TELEFAX: (619) 235-0176
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 896 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-480-145-2

Query Match 7.8%; Score 486.5; DB 1; Length 896;
 Best Local Similarity 20.5%; Pred. No. 4.3e-26;
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;
 15 WAITEEBAKHDOQLSLKPIAGFITGDARVFFQSGLPQVLAQIVMALDMNNDGMD 74

Db 122 WAVEPEDKATYDAIFDLSLVNGLSGDYKVPYLANSKLVLDLGRWELSDIDHGMLD 181
 Qy 75 QVEESIMKLIKLLQGYQLPSTLPVWKQOPVAISSAPFGIGIASMPPLTVAAYPM 134
 Db 182 RDEFAVAMFLVYCALE-----KEEVPFM 203
 Qy 135 GSITVGMSPPLVSSVPAAPVPLANGAPVIOPLPAFAHAPATWPKSSPSRSGPSOL 194
 Db 204 -----SLPPALVPP-----SKR 215
 Qy 195 NTKLOKAQSPDVAAPAAWAVPQSSRLKYROLFNSHDTWSGHLTGPOARTIMQSSL 254
 Db 216 KT-----WVVSAPAKAYDEFLTDMQMGVFGLEVRBEFLTGTG 257
 Qy 255 PQOGLASTWNSLIDQCKLTAEFFILAMHLIDVAM-SGQPLPVLPPEVTPSFRVRS 313
 Db 258 PSTLAAHWISLIDTKDCKLSKQFALAFHLSQKLKGIDPPHVLTPPEMIPPS--DRA 314
 Qy 314 GSGMSVSSSSVDQRLPEBPSSEDEQPEKKLPVTFEDKKRENERGSELEKROALLE 373
 Db 315 SLOKNITIGSSPV-----ADFSAIK-----ELDTIANETVD 344
 Qy 374 QQRKEORLAQLERAQERKERERQEQAKROLELEKOLEQOREERREKLEIR 433
 Db 345 LQREK-----NVVEQDLKEKE-----DTIKORTSEVQ--DLQDEVORE 380
 Qy 434 EAARBELEOROLEMENRROELINQNKKEOGTVLKARKTLEFLEALNDKHOLEG 493
 Db 381 NTNLQKLOAQKQ-----QVQELDE-----LDEQKQLEB 410
 Qy 494 KLDIRCRLATQROEISTNKSRELIAETHLQOOLQOESQWGLRILPEKQILSDQK 553
 Db 411 QLKVEKKCAEAOILSSLK-----AELTQESQSTYBEBELAK-----ARELSR 456
 Qy 554 VQONSRLRDSLLTLKALAEKELARQOLREQLREVERETS-----KLOEIDVFNQK 607
 Db 457 LQGETAE-----LEESVESKQALEPQOHLQDSQOEISMQWKLMEKMLEHNSQLN 510
 Qy 608 ELREIHS-----KOQLOKQSLAEARKQOEKESLEL----- 641
 Db 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVQGSNLESEPTHOESPASPELLPSG 570
 Qy 642 EKOKEDAQRYQOEEDKQMLEHVQOEBQRPKPKPEERLKRKEDSVRKKAEER----- 694
 Db 571 VTDENEVTTATYKVCSELDN-----NRHSKEEDPFNVDSSTLGPVADTULDFPQS 622
 Qy 695 -----AKPEMODKOSRLFHPHOEPK-----LATOAPMSTTEKGLTISAQESV 738
 Db 623 DPFVGSDFPKDDPFCKIDPFQGDPPKSDPPASDCFFRQSTDPFATSTSDPFAANSSI 682
 Qy 739 KVVYRRALYPFESRSHDEITIQPGDIYVWDESQT-----GEPMGLG----- 780
 Db 683 TSV-----ETLKNDPFAFGTVAASDASIDPPASVFGNSFGGFPADFSTLSKVN 734
 Qy 781 ---ELKGTGMFPAN-----YAEKIPENEVPTPAKPVTDLTSAAPKALRETAPLP 830
 Db 735 NEDPRSATSSSVSNVITKAVFETSVKSEDEPALP-----PKIGTPRCPPLP 785
 Qy 831 -----VTSSEP-----STTPNNMADPSSSTWSSSNEKPEFTDMDT 865
 Db 786 PGKSINKLDSPPDKLNDPFPQPPGNDSPKXOPFEMFCDFTSATTTNKADPSPNFAN 845
 Qy 866 WAAOPS 871
 Db 846 FSAIYS 851

RESULT 6
 US-08-477-389-2
 ; Sequence 2, Application US/08477389
 ; Patent No. 5872219
 ; GENERAL INFORMATION:

```

APPLICANT: DiFiore, Pier P
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
CITY: Newport Center Drive, Sixteenth Floor
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 896 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-389-2

```

```

Query Match 7.8%; Score 486.5; DB 2; Length 896;
Best Local Similarity 20.5%; Pred. No. 4.3e-26;
Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

```

```

15 WAITVEERAKHQDQFLSTKPIAGFTIGDQARNFPQSGLPQVLAQIMALAMNNNGRD 74
122 WAVEKEDRAKKAIFGSLSPVNGFLSGDVKPVVLNSKLPVDLGRWELSDHDHGMID 181
75 QVEFSIAMLKIKLQGYQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLAAVAPVM 134
182 RDEPAVAMFLVYCALE-----KRPVPM 203
135 GSIPVVGMSPLVSVPPAAVPPPLANGAPVVIQPLPAFAHPAATWPKSSFSRSGGSQL 194
204 -----SLPPLVPP-----SKR 215
195 NTKLQKAGSFDVASAPPAEMVPOSSRLKYROLFNSHDKTWSGHLTGPQARTIMQSSL 254
216 KT-----WVSPAEKAKTDEIFLKTDKMDGFGVSGLEVEIFLKTGL 257
255 PQAQIASIWNLSDDIDQGLTAEEFLIAMHLIDVAM-SCQPLPVPVLPPEYIPPSFRFVRS 313
258 PFTLLAHIMSLCTDKQCGKLSKQGFALAPHLISQKLIKIGIDPPHVLTPMIPPS---DRA 314
314 GSGMSVISSSVQRLPEPSSSEDEQPKLPTVFEDKKRENFERSVLEKRRQALIE 373
315 SLQKNIIGSSPV-----ADPSAK-----ELDTLNNEIYD 344
374 QQRKEQERLAQLEBAQERKRRERQEAQKOLEKQLEKORERLEKREERKXIERR 433
345 LQREK-----NVEQDLKEKE-----DTTKQRTSVQ---DLQDEVORE 380
434 EAKKELEERQOLEWERNRROELANDNRNKEOGTVVLKARKRTLEFELBALNDKKHLEG 493

```

```

DB 381 NTKLQKQAKQ-----QVCELDE-----LDEQKQLE 410
494 KQDJRCRLATQROBIESTNKRELRIRAEITHLQOOLQESQOMGLRIPKQILSQKQ 553
411 QKEVRKKCAEAQILSLK-----AEVTSQESQISTYEBELAK-----ABEELR 456
554 VQONSLHRDLSLTLRALAEKALAQOLREQDEVERETR-----KLQEDIVFNQK 607
457 LQQETAE-----LEESVESGAQLEPLQCHQDSQQLSSQMKLMEMKDLNNNSQLN 510
608 EUREIHS-----KQOLQKRSLEAARLKQKEQERSLEL--- 641
511 WCSSPHSILVNGATDYCSLSTSSSETAMINEHEVQCSNLESEPIHQESPARSSPELPG 570
642 EKQXDAQRQVQERKQMLHENVQOEQRPRKPHEDRLKEDSVRKKEAER----- 694
571 VTDENEVTTAVTEKCSIEDN-----NRHSKEEDPPNVSSSLTGFAVATNLDPFOS 622
695 -----AKPEMOKQGRLEFHPQEPAP-----LATQAPWSTTEKPTLISQESV 738
623 DPFVSGDPPKDDPPKIDPFQDPPKSGDPPASDCFRQSTDPFATSSDTPFSAANSSI 682
739 KVVYRALYPFESRSHDEITTPGDIWVDESQT-----GEPGMLG----- 780
683 TSV-----ETLKHNDPFAFGTVVAASDSDATDPFASVFGNESFGGFADPSTLSKVN 734
781-----ELKKTGMFPAN-----YAKIPENEVPTAKPPTDLSAPAPLARETPAP 830
735 NEDPFRSATSSSVSNVITKNVFEETSVKSEDEPPALP-----PIGTPTRCP 785
831 -----VTSEEP-----STPNNMADFSSTWSSNEKEPETNMOT 865
786 PKGRSLNKLDSPDKLNDPPQPFPGNDSPKKDEPMFCDPFTSATTTINKADPSNFPAN 845
866 WAAQPS 871
846 FSAYPS 851

```

```

RESULT 7
US-08-095-737-4
Sequence 4, Application US/08095737
Patent No. 5487979
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
CITY: Newport Center Drive, Sixteenth Floor
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:

```


SEQUENCE CHARACTERISTICS:
 LENGTH: 897 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-095-737-4

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.7e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WAITVERAKHDOQFLSLKPIAGITGDOANFFQSGLPQVLAQIWAADMMNDGMD 74
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 122 WAVESEDAKAYDAFDLSLSPDGLSGKVPVLNLSKLPVILIGRWELSDIDHDKLD 181
QY 75 QVESIAMIKLKILKIQGYQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLTAVAPVM 134
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 182 RDEPAVAMFLVYCALE-----KEPYPM 203
QY 135 GSIPVGMSPPLVSSVPPAAPPVLANGAPVIOPLPAFAHPAATWPKSSFSRSGPSQL 194
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 204 -----SLPPALVPP-----SKR 215
QY 195 NTKLQKQSFVVASAPPAEWAVPQSSRLKYROLFNHDKTMSGHLTGPOARTILMOSSL 254
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 216 KT-----WVSPAERAKAYDEIFLKTDKMDGVSGLEVRERETLTKGL 257
QY 255 POALASIMWLSIDIDQDKLTAEFFILAMHLIDVAM-SGQPLPVLPEYIPPSFRVRS 313
    ||:|||||:|||||:|||||:|||||:|||||:
DB 258 PSALLAHIMWISCDTFCGCKLTKSKDQFALAFHLINQKIDPPHSLTPKEMIPPS---DRS 314
QY 314 GSGMSVSSSSVDQRLPEPSSSEDOQPEKKLPVTFEDKKENPERGSEVELEKRRQALLE 373
    ||:|||||:|||||:|||||:|||||:|||||:
DB 315 SLQKNITGSSPV-----ADPSAIK-----ELDTLNNEITVD 344
QY 374 QQRKEOERLAOLEPQERKERERQEQEAKQLELEKOLEKQERELERQEEERKEIERR 433
    ||:|||||:|||||:|||||:|||||:|||||:
DB 345 LQREK-----NNEQDLKEKE-----DTVQRQISEVQ---DLQDEYQRE 380
QY 434 EAARKELEERQOLEWERNRROELLNQRNKEQEGTVLAKRRKLTFFELBALNDKKHOLEG 493
    ||:|||||:|||||:|||||:|||||:|||||:
DB 381 SINQKIQAKQK-----QVQELLGE-----LDEQKQALEE 410
QY 494 KLQDIRCRLATQROEISTNKSRLRIETALHLOQLQESQOMGLRLEPKQILSDQKQ 553
    ||:|||||:|||||:|||||:|||||:|||||:
DB 411 QLOEVRKKAEEAQLISLK-----AEITQSQISSEYEEELK-----ARELSR 456
QY 554 VQONSILHSDLSLTLKRALEAKELARQQLREQLDEVERE-----TRSKLQEIIVENNOLKE 608
    ||:|||||:|||||:|||||:|||||:|||||:
DB 457 LQQETAQ-----LEESVESGKAQLEPLQCHLQESQOQIISSMQKRLKMDJETNNQ--- 507
QY 609 LREIHSKQQLQKQ-----RSLEAPARKQKEQERKSLLEKQKEDQAQRVQ 653
    ||:|||||:|||||:|||||:|||||:|||||:
DB 508 -----SNWSSSPQSVLVNATDYCSLSTSSSETANFNHAGQNNLESEPHQSSSVRS 562
QY 654 ER-----DKQWLEHVQGEQEPQPRKHEEDRLKREDSVAKKEA-----EERAKPEMDK 702
    ||:|||||:|||||:|||||:|||||:|||||:
DB 563 PEIAPSDVTDESEAVTAVAGNEKVTPR--FDDDKISKKEEDPPNVSSSLTDVAATNDLDF 620
QY 703 QSRLF-----HPHOEPAKLATQAWSTTEKGPLTISQAESQKVVY 743
    ||:|||||:|||||:|||||:|||||:|||||:
DB 621 QSDPFGVSDPFPKIDPFGGDPFPKSDPFPASDCFFKQTSIDPFTTSSDTPFSASNN 680
QY 744 RALYPFESRSHDEITIQGDIVMWDESQTEP--GWLGLGKGTGMFPANVYAEKIPENE 801
    ||:|||||:|||||:|||||:|||||:|||||:
DB 681 SSNTSVETWTKNDPFAFGTIVAAASDQATDPFASVFQNESFG--DGFAFDSTLSKVNED 739
QY 802 VPTP-----AKVTDLTGA-----PA--PKALARETAPLPVTSSESTPNN 842
    ||:|||||:|||||:|||||:|||||:|||||:
DB 740 AFNFTIISSTSSVTIAKPMLEETASKSDVPPALPKVGTPTRPCPP-----EQKRIN 794
QY 843 WADPSSST-----WPSSSNEKPEPTDNDMTAAQPSLTVPSAGQLRQSAFTPATATGSS 895
    ||:|||||:|||||:|||||:|||||:|||||:

```

DB 795 KLDSDDLKLNDFPQPPGNDSPKEXQPDWFC-----DPFTSSTTNKE 838
 QY 896 PSP 898
 DB 839 ADP 841

RESULT 8
 US-08-480-145-4
 Sequence 4, Application US/08480145
 Patent No. 5717067

GENERAL INFORMATION:
 APPLICANT: Diflore, Pier P
 APPLICANT: Razioli, Francesca
 TITLE OF INVENTION: A Substrate for the Epidermal Growth
 FACTOR RECEPTOR KINASE
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson & Bear
 STREET: 620 Newport Center Drive, Sixteenth Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: United States of America
 ZIP: 92660

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/480,145
 FILING DATE: 07-JUN-1995

CLASSIFICATION: 530
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/095,737
 FILING DATE: 22-JUL-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelsen, Ned A

REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH060.001A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:

LENGTH: 897 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-480-145-4

Query Match 7.8%; Score 486; DB 1; Length 897;
 Best Local Similarity 21.3%; Pred. No. 4.7e-26;
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

QY 15 WAITVERAKHDOQFLSLKPIAGITGDOANFFQSGLPQVLAQIWAADMMNDGMD 74
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 122 WAVESEDAKAYDAFDLSLSPDGLSGKVPVLNLSKLPVILIGRWELSDIDHDKLD 181
QY 75 QVESIAMIKLKILKIQGYQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLTAVAPVM 134
    ||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 182 RDEPAVAMFLVYCALE-----KEPYPM 203
QY 135 GSIPVGMSPPLVSSVPPAAPPVLANGAPVIOPLPAFAHPAATWPKSSFSRSGPSQL 194
    ||:|||||:|||||:|||||:|||||:|||||:
DB 204 -----SLPPALVPP-----SKR 215
QY 195 NTKLQKQSFVVASAPPAEWAVPQSSRLKYROLFNHDKTMSGHLTGPOARTILMOSSL 254
    ||:|||||:|||||:|||||:|||||:|||||:
DB 216 KT-----WVSPAERAKAYDEIFLKTDKMDGVSGLEVRERETLTKGL 257
QY 255 POALASIMWLSIDIDQDKLTAEFFILAMHLIDVAM-SGQPLPVLPEYIPPSFRVRS 313
    ||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 258 PSALLAHWSLCTKCGKSLKQDQFALAFHLINQKIGIDPESHITPEMIPPS---DRS 314
Qy 314 GSGMSVSSSVQDRLPEERSSSEDOPEKCLPTEDEKRENFERSVELEKROALLE 373
Db 315 SLQKNITGSSPV-----ADSSAIK-----ELDTLANEIVD 344
Qy 374 OORKEOERLAQERAEERKEREREOEAKROLELEKOLEKORELEROREERKEIER 433
Db 345 LOREK-----NNVQDLKEKE-----DTVKQRTSEVQ---DLQDEVOKE 380
Qy 434 EAKRELEROROLEWERNRROELNQRNKEQEGTVLAKARKTLEFELEALNDKQKOLEG 493
Db 381 SINLQLOAQOKQ-----QVQELIGE-----LDQKQALE 410
Qy 494 KLQDIRCRATQROEIESTNKSRELIAETIHLQOQLOESQOMLGRLPEKOLISQLOK 553
Db 411 QLOEVAKKAEKQALISSLK-----AETQESQISSYEBELK-----AREELR 456
Qy 554 VOONSILHRDLSLTKRALEKELARQOLREOLDEVERE-----TRSKLOEIVFNQKLE 608
Db 457 LOQETMQ-----LESVESGKQLEPLQOHLQESQOELISMQMRLEMKDLETNNQ--- 507
Qy 609 LREIHSKQOLOKQ-----RSLFAARLKQKQERKSLDEKQEKEDAQORVQ 653
Db 508 -----SNMSSSPQSVLVNGATDYCSLSTSSSETANFNEHAGQNNLESEPTHQESSVRSS 562
Qy 654 ER-----DKQMLEHVQOEOPRPKPKHEEDRLKREDSVRKKEA-----EERAKPEMQDK 702
Db 563 PEIAPSDVDESEAVTVAGNEKVTPR--FDDDKHSKEEDPFNVESSSLTDAVADTNLDPF 620
Qy 703 QSRLE-----HPOEPAKLATQAPWSTTEKGLTISAQSVKVVVY 743
Db 621 QSDPFQSGDPFKDDPFGKIDPFGCDPFKSGSDPFASCFKQISTDPTTSSITDPFASASN 680
Qy 744 RALYPESHSHDEITIQGDIVWVDSQTEGP--GWLGGELKQKTMFPANVAEKIPENE 801
Db 681 SNTSVETWKHNDPFAFGTVVAASDASATDPFASVFGNESFG--DGFADPSTLSTKNNED 739
Qy 802 VFRP-----AKPTDLTSA-----PA--PKLAIRETPALPVTSSSEPTTPNN 842
Db 740 AENPTISSSTSVTLKPMLEETASKSEBVPALPKVGPTRPCPP-----FGKRPAIN 794
Qy 843 WADFSST-----WPSSNEKPEETDNPMTMAQPSLTVPSAQQLAQRSAFAPATAGSS 895
Db 795 KLDSSPLKLNDFPQDFPQNDSPKXKOPDMFC-----DPTSSITTNKE 838
Qy 896 PSP 898
Db 839 ADP 841

```

RESULT 9
 US-08-477-389-4
 / Sequence 4, Application US/08477389
 / Patent No. 5872219
 / GENERAL INFORMATION:
 / APPLICANT: DiFiore, Pier P
 / TITLE OF INVENTION: A Substrate for the Epidermal Growth
 / FACTOR RECEPTOR KINASE
 / NUMBER OF SEQUENCES: 4
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Knobbe, Martens, Olson & Bear
 / STREET: 620 Newport Center Drive, Sixteenth Floor
 / CITY: Newport Beach
 / STATE: California
 / COUNTRY: United States of America
 / ZIP: 92660
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: PatentIn Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-389-4

Query Match 7.8%; Score 486; DB 2; Length 897;
Best Local Similarity 21.3%; Pred. No. 4,7e-26;
Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

15 WAITVERAKHQOFLSKPIGFTTGQARNFFQSGLPQVLAQIYALADMMNDGRND 74
122 WAKSEDKAKYPAIFDPSLPVQDGLSKPVLNLSKLPVILGRWELSDIDHGKLD 181
75 QVEPSIAMLILKQGYLPSTLPVVMQGVVAISSAFAFGIAGIAMPPLTAVAPVM 134
182 ROEFAPMPLVYCALE-----KEPVPP 203
135 GSIPVVGMSPLVSSVPAVPEPLANGAPVIOPLPAFAHPATWPKSSFSRSGPSQI 194
204 -----SLPPLVPP-----SKR 215
195 NTKLQAKQSFVASAPPALEMAVPOSSRLKYROLFNSHDKTWSGHITGQARTILMQSSL 254
216 KTL-----WVSPAEKAKYDEIFLTKDKMDGVYSGLEVETFLTKGL 257
255 POAQLASINLSIDIDQGLTAEFTLAWHLIDVAM--SQGPLPVLPPEYIPSPFRVRS 313
258 PSALLAHWSLCTKCGKSLKQDQFALAFHLINQKIGIDPESHITPEMIPPS---DRS 314
314 GSGMSVSSSVQDRLPEERSSSEDOPEKCLPTEDEKRENFERSVELEKROALLE 373
315 SLQKNITGSSPV-----ADSSAIK-----ELDTLANEIVD 344
374 OORKEOERLAQERAEERKEREREOEAKROLELEKOLEKORELEROREERKEIER 433
345 LOREK-----NNVQDLKEKE-----DTVKQRTSEVQ---DLQDEVOKE 380
434 EAKRELEROROLEWERNRROELNQRNKEQEGTVLAKARKTLEFELEALNDKQKOLEG 493
381 SINLQLOAQOKQ-----QVQELIGE-----LDQKQALE 410
494 KLQDIRCRATQROEIESTNKSRELIAETIHLQOQLOESQOMLGRLPEKOLISQLOK 553
411 QLOEVAKKAEKQALISSLK-----AETQESQISSYEBELK-----AREELR 456
554 VOONSILHRDLSLTKRALEKELARQOLREOLDEVERE-----TRSKLOEIVFNQKLE 608
457 LOQETMQ-----LESVESGKQLEPLQOHLQESQOELISMQMRLEMKDLETNNQ--- 507
609 LREIHSKQOLOKQ-----RSLFAARLKQKQERKSLDEKQEKEDAQORVQ 653
508 -----SNMSSSPQSVLVNGATDYCSLSTSSSETANFNEHAGQNNLESEPTHQESSVRSS 562
654 ER-----DKQMLEHVQOEOPRPKPKHEEDRLKREDSVRKKEA-----EERAKPEMQDK 702
563 PEIAPSDVDESEAVTVAGNEKVTPR--FDDDKHSKEEDPFNVESSSLTDAVADTNLDPF 620

```

```

QY OSRL-----HPHOBPAKLTQAPMSTTEKGLTISAOSBKVYX 743
Db QSDPFVSGDPFKDDPFGKIDPFPGGDPFKSDPFPAASCFFKQHTDFTTSSIDPFASBN 660
QY 744 RALYPFESRSHDEITIQPGDIYMVDESQTGP--GWLGGELKGTGMFPYANAEKIPENE 801
Db 681 SSNTSEVETWKHNDPFAFGCTVVAASDSATDPPASVFGNESFG--DGFADPFTSLKYNED 739
QY 802 VPT-----AKPTDLTSA-----PA--PKLARETPAFLPVTSSEPTTPNN 842
Db 740 AFNPTSSSTSSVTTAKPFLLEETASGEVFPALPFPKVGTPTRPCPP-----PGKRPIN 794
QY 843 WADFSST-----WPSSNEKEPETHNDMTWAQSLTVPSAGLQRORSAFTPATIGSS 895
Db 795 KLDSSDPLKNDPFPQFPFGNDSFKEDPDMFC-----DPFTSTTTNKE 838
QY 896 PSP 898
Db 839 ADP 841

```

RESULT 10
 US-08-728-323A-2
 : Sequence 2, Application US/08728323A
 : Patent No. 5948676
 : GENERAL INFORMATION:
 : APPLICANT: Chang, Yuan
 : APPLICANT: Bohenzky, Roy A.
 : APPLICANT: Russo, James J.
 : APPLICANT: Edelman, Isidore S.
 : APPLICANT: Moore, Patrick S.
 : TITLE OF INVENTION: Immediate Early Protein From Kaposi's
 : TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
 : TITLE OF INVENTION: Encoding Same And Uses Thereof
 : NUMBER OF SEQUENCES: 21
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10036
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent In Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/728,323A
 : FILING DATE:
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: White, John P.
 : REGISTRATION NUMBER: 28,678
 : REFERENCE/DOCKET NUMBER: 0575/52268/JPW/NSC/SKS
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 212-278-0400
 : TELEFAX: 212-391-0525
 : INFORMATION FOR SEQ ID NO: 2:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1162 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-728-323A-2

Query Match	6.1%	Score 382	DB 2	Length 1162
Best Local Similarity	24.6%	Pred. No. 1.6e-18		
Matches 127	Conservative 139	Pid. Matches 197	Indels 54	Gaps 19
QY	327	ORLP--PPSSDEDO--PEKTLPTFEKRENERGSEVLEKRRALLTEQRKEQERTL	382	
	:		:	
	:		:	

[illegible]

```

RESULT 11
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
; APPLICANT: Kieff, Elliott D.
; APPLICANT: Ballestas, Mary E.
; APPLICANT: Kaye, Kenneth M.
; TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/298,568
; CURRENT FILING DATE: 1999-04-21
; EARLIER APPLICATION NUMBER: US 60/109,422
; EARLIER FILING DATE: 1998-11-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

```

```

Query Match 6.1%; Score 382; DB 4; Length 1162;
Best Local Similarity 24.6%; Pred. No. 16-18;
Matches 127; Conservative 139; Mismatches 197; Indels 54; Gaps 197;

QY 327 QRLPE--EPSSDEBQ--PEKKLVPTEDKKENFERSGVLEKRRQALLQQRKEQRL 382
Db 551 QRPQQRPEPQQRPEPQQRPEPQQRPEPQQRPEPQQRPEPQQRPEPQQRPEPQQR 608
QY 383 AQLERAQERKRRKROQEAQRQLELQLEKQLELRQREERKKELEIRERBAKRELEIR 442
Db 609 EQ-QQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDEQQQDE--QQQDE 664
QY 443 QRLERMRNRQELLNRKNEQDEGTVLKARKTLFEFLMLNPKQQLBKLDQIRRL 502
Db 665 QQQDEQQQDEQQQDEQQQDEQQQDE--EQQDEQQQDEQQQDEQQQDEQQQ 709

```

```

0Y  ATOPOEISTNKSREIRAIETHLOQOLOSOMGLRLPKQJLSLDOLKYOQNSLHND 562
Db  DEQEBOQDEQOQDEQOQOQDEQOQOQDEQOQOQDE--QOQOQDEQEBOQEBOQE 767
0Y  SLLTLKRLAEAKELARQOOLREOIDEVERETTSLOEIDVPNNOLKELREHNSQOQOKO- 621
Db  ELEBOQEHEDEQELEBOQEBOQEBOQEBOQEBOGLE--BOQGLE--QEQELEBOE 822
0Y  RSLF--AARLKQOEOKRSLELEKOKEDAORRVOERDKOMLEHVOQEORPRPKPHE- D 678
Db  QELEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQEBOQE 882
0Y  RLKREDSVRKKAERPAKPEMOD---KOSRLFHHPQERPAKATQAPMSITTEKGLPLTI-- 732
Db  EYEBEOBOELEVEEBOEBOELEVEEBOEBOELEVEEBOEQOQGEVQOEOETVEE--PILHG 941
0Y  -SAGSIVVYVYRALLYPFSRSHDEI-TIQOPDIWVNDSSQGEFPMGLGELKQKQWPR 790
Db  SSSSEMEEMVD-----YPVVS-THEQJASSPPGQNTPEDDDPQPSREYRVVLITSPPHRR 995
0Y  ANYAEKIPENEVYTPAKPVTDLTSAAPAKLTALRETPA 827
Db  GYMRKRVP--VTHPKKPHRYQOPRPVPRQJDDCPA 1029

```

```

RESULT 12
US-09-410-399-2
: Sequence 2, Application US/09410399
: Patent No. 6482587
: GENERAL INFORMATION:
: APPLICANT: Robertson, Erle S.
: APPLICANT: Cotter, Murray A.
: TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
: TITLE OF INVENTION: to Genomic Host DNA
: FILE REFERENCE: UM-03778
: CURRENT APPLICATION NUMBER: US/09/410,399
: CURRENT FILING DATE: 1999-10-01
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 1162
: TYPE: PRT
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

```

[illegible]

```

Qy 679 RLRKREDSVAKKEAEBAERAKEMQD-----KQSLFLFHQBPALALQAQWSTTEKCEPLI-- 7322
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 883 EYVEQGEQLEEEVEEEOBEQEBELEVEEEOQBLEEEVEEEOGVQBOEQETVEE--PIILHG 941HG
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 733 -SAQESVKKVYRYRALYPPFSRSDEI-LTIQGDIVMVEDSQTGEKPMGLGELKCKTKTMFP 790
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 942 SSSEDEMEVD-----YPVVS-THEQIASPQGDNTPDQDPQGPSREYRYVALRTSPHP 995
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Qy 791 ANYAERKIPENBEVTPAKPVTDLTSAAPAKLALRETPA 827
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 996 GVMARRRP---VTHPKKHPRIQDEPVDYRQIDDCPA 1029
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 13
US-08-056-200-94
Sequence 94, Application US/08056200
Patent No. 5616500
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: "Ricinoyalin and Transglutaminase-3 and
METHOD OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESSES:

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

	Query Match	6.1%; Score 381; DB 1; Length 1898;
	Best Local Similarity	27.0%; Pred. No. 3.9e-19;
	Matches	136; Conservative 117; Mismatches 15; Indels 96; Gaps 21
OY	326 DQRLPEBSSDEDOPEKQLPVTFEDDKRNFENRGSVLEKRQALLBEQ--RKQE--R	381
	: : : : : : : : : : : : : : : : :	
Db	348 EQQLRR---QEERRRQQQLRRQEERREQQQLRRQEERREQQQLRRRQQQLRR	404
OY	382 LAQLRAEQEKERE-RQEQAKR-QLELKQLEKQELEROREER-----	426
	: : : : : : : : : : : : : : : : :	
Db	405 EQQLRRQQQLRRQQQLRRQQQLRRQQQLRRQQQLRRQEERREHKEGHEQRREQ	464
OY	427 ---RKELRRRAARGLERQRQLEMERNRQELLNQRKQEGDGVVLAARRITLFEFLA	483
	: : : : : : : : : : : : : : : : :	
Db	465 RLKKEOERRRWMLKREETERH---EQRRRQOL-KRQEEB-----RRRWMLTEEB-	513

Query Match 6.1%; Score 381; DB 2; Length 1898;
 Best Local Similarity 27.0%; Pred. No. 3,9e-18;
 Matches 136; Conservativity 117; Mismatches 154; Indels 96; Gaps 21;

484 LNDKKGLEGLQDIDRCRLATQROEISTNKSRELIAITHLQOOLQESQ-----OML- 537
 514 --ERREOQERREOQIR-REOERREOQLKREOERREOQLRSEOOQLREOERREOQLK 570
 538 -----GRLIPEKQILSDOL-----KQVOQNSLHRDLSLTLKRALFAKELARQ 580
 571 REEKRLEQERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 630
 581 LREOQLDE-VERETRSKLOEIDVFNNQLELR---EHSKQOLQKORSL---EAAFLKOK 632
 631 QOEERDERLKRREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 690
 633 -----EOKRSLLEI-EKQKEDQORRVOERDKOMLEHVQOEBQPRPK----- 673
 691 LKREHEERREOQLREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 750
 674 PHEEDRLKREDSVAKKEAEERAKPEMODKQSLRFLPHQEPKATLQAPWSTTEKGPLTIS 733
 751 PQOQEKRRRESELQWOQEEERAHQOQOEEERDFTWQOAE-----EKSERGRQRLS 803
 QY 734 AQESVAVVYVYALYPFESRSHDE 756
 DB 804 AAPPLREORERQLRAEERQOREQ 826

RESULT 14
 US-08-800-644-94
 ; Sequence 94, Application US/08800644
 ; Patent No. 5958752
 ; GENERAL INFORMATION:
 ; APPLICANT: Steiner, Peter M.
 ; APPLICANT: Lee, Seung-chul
 ; APPLICANT: Kim, In-Gyu
 ; APPLICANT: Chung, Soo-Il
 ; APPLICANT: Park, Sang-Chul
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
 ; TITLE OF INVENTION: Methods of Using Same
 ; NUMBER OF SEQUENCES: 117
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive, Sixteenth Floor
 ; CITY: Newport Beach
 ; STATE: CA
 ; COUNTRY: U.S.A.
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/800,644
 ; FILING DATE: 14-FEB-1997
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION NUMBER: US 08/056,200
 ; FILING DATE: 30-APR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fedick, Michael F.
 ; REGISTRATION NUMBER: 36,799
 ; REFERENCE/DOCKET NUMBER: NIH054,001A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (714) 760-0404
 ; TELEFAX: (714) 760-9502
 ; INFORMATION FOR SEQ ID NO: 94:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1898 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-800-644-94

Query Match 6.1%; Score 381; DB 2; Length 1898;
 Best Local Similarity 27.0%; Pred. No. 3,9e-18;
 Matches 136; Conservativity 117; Mismatches 154; Indels 96; Gaps 21;

QY 326 DQRIPEPSSDEOQPKLPTFEDKRENFGSVLEKRAQALIEQ--RKEQ--R 381
 DB 348 EQQLRRE---QEERREOQLRREOERREOQLRREOERREOQLRREOERREOQLRREOERREOQLR 404
 QY 382 LAQLEAEORKEE--ROEBAKR--OLELEKOLEKRELEORREER----- 426
 DB 405 EQQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLRREOQLR 464
 QY 427 ---RKEIEREAAKRELEOROLEWERNRROELINQNKQOEGTVVAKARKTLEFLEA 483
 DB 465 RKRQOERREOQLRREOERREOQLRREOERREOQLRREOERREOQLRREOERREOQLRREOERREOQLR 513
 QY 484 LNDKKGLEGLQDIDRCRLATQROEISTNKSRELIAITHLQOOLQESQ-----OML- 537
 DB 514 --ERREOQERREOQIR-REOERREOQLKREOERREOQLRSEOOQLRSEOOQLRSEOOQLRSEOOQLR 570
 QY 538 -----GRLIPEKQILSDOL-----KQVOQNSLHRDLSLTLKRALFAKELARQ 580
 DB 571 REEKRLEQERREOQLRREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 630
 QY 581 LREOQLDE-VERETRSKLOEIDVFNNQLELR---EHSKQOLQKORSL---EAAFLKOK 632
 DB 631 QOEERDERLKRREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 690
 QY 633 -----EOKRSLLEI-EKQKEDQORRVOERDKOMLEHVQOEBQPRPK----- 673
 DB 691 LKREHEERREOQLREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLKREOERREOQLK 750
 QY 674 PHEEDRLKREDSVAKKEAEERAKPEMODKQSLRFLPHQEPKATLQAPWSTTEKGPLTIS 733
 DB 751 PQOQEKRRRESELQWOQEEERAHQOQOEEERDFTWQOAE-----EKSERGRQRLS 803
 QY 734 AQESVAVVYVYALYPFESRSHDE 756
 DB 804 AAPPLREORERQLRAEERQOREQ 826

RESULT 15
 US-08-475-894-2
 ; Sequence 2, Application US/08475894
 ; Patent No. 5641748
 ; GENERAL INFORMATION:
 ; APPLICANT: Yen-Ming Hsu
 ; TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, Suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/475,894
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Louis Myers
 ; REGISTRATION NUMBER: 35,965
 ; REFERENCE/DOCKET NUMBER: BGP-191
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 227-5941
 ; INFORMATION FOR SEQ ID NO: 2:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: 1linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-475-894-2

```

```

Query Match          5.3%; Score 333; DB 1; Length 553;
Best Local Similarity 27.4%; Pred. No. 1,7e-15;
Matches 93; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

```

```

QY 912 QALYPMRAKKNHNLNFKSDVITVLEQD-MMMFGEVQOGKMFPSYVKLISGPYRKST 970
DB 77 QDIYPLYLTHDDELITISVGEIITNIRKEDGXWMEGQINGRGLFPDNFVREIKEMKXP 136
QY 971 SIDTGPTESPASIKRVASPAKPAIPGEPT-----AMTYESESQGLTF 1016
DB 137 LTNKAP-EKP--LHEV--PSGNSLSSEFILTNRKGERRRRRRCQVAFSYLPQNDDEL 191
QY 1017 QOGDVI-VYTKKGDPMWGTGVDKSGVFPSNYVR-----LKDSE 1054
DB 192 KVGDIIEVVGVEEGWEGVNGKTMFSPNFIKELSGSDEIGISQDEQLSKSLRETT 251
QY 1055 GSGTAGKTGSLGKKEPIAQVIASVATGPEOL-----TLAPGQLILIRKKNPGWMEGE 1108
DB 252 GSESDEGDSS-STKSGANGVATATAIQPKYKGVGFQDIFKDKPIKLRPS-----IYVE 306
QY 1109 LQARGKROIG-WFPANYVKLISPGTSKITPTLPTFAVQPAVCOVIGMYDYTAQNDEL 1167
DB 307 NDPLPEKRTIGKKLPAAT--TATPDSSK--TEMDSRTSKDYCKVI--FPYEAQNDDEL 358
QY 1168 AFSKGOIIVLNKE--DPDMWKEVSGOVLFPSPNYVKL 1204
DB 359 TIKEGDIVTLINNDICIVGMEGELNGRGRGVFPDNFVKL 397

```

```

Search completed: December 4, 2003, 15:21:35
Job time : 28 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 15:20:14 ; Search time 34 Seconds

(without alignments)
6635.241 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFPTPGGSLDVAITVE.....VGLPFSNYKLTMDPSQ 1213

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058.5	32.8	462	9	US-09-879-957-38
2	1388.5	22.1	509	9	US-09-879-957-194
3	1330.5	21.2	530	10	US-09-764-868-78
4	816.5	13.0	248	9	US-09-879-957-40
5	572	9.1	144	11	US-09-764-881-13
6	527	8.4	813	10	US-09-964-899-25
7	431.5	6.9	224	11	US-09-764-881-112
8	429	6.8	90	9	US-09-864-761-33620
9	429	6.8	90	9	US-09-864-761-43004
10	426	6.8	90	9	US-09-864-761-33635
11	422	6.7	82	9	US-09-864-761-46565
12	413	6.6	87	9	US-09-864-761-34169
13	409.5	6.5	168	11	US-09-764-881-128
14	382	6.1	1162	12	US-10-294-804-2
15	363.5	5.8	2701	15	US-10-171-311-83

16	343.5	5.5	843	12	US-10-144-194A-84	Sequence 84, Appl
17	333.5	5.3	779	12	US-10-144-194A-86	Sequence 86, Appl
18	330	5.2	1239	11	US-09-291-417-13	Sequence 13, Appl
19	325	5.2	659	14	US-10-144-621-2	Sequence 2, Appl
20	322	5.1	59	9	US-09-879-957-133	Sequence 133, App
21	319	5.1	735	12	US-10-144-194A-109	Sequence 109, App
22	319	5.1	784	12	US-10-144-194A-108	Sequence 108, App
23	316.5	5.0	287	10	US-09-764-868-744	Sequence 744, App
24	315.5	5.0	1170	12	US-10-341-434-95	Sequence 95, Appl
25	315.5	5.0	1564	12	US-10-144-198-2	Sequence 2, Appl
26	315.5	5.0	1564	12	US-10-144-198-4	Sequence 4, Appl
27	314	5.0	56	9	US-09-879-957-135	Sequence 135, App
28	309	4.9	62	9	US-09-879-957-134	Sequence 134, App
29	304.5	4.9	811	15	US-10-128-714-3314	Sequence 3314, Ap
30	303.5	4.8	611	9	US-09-216-393-81	Sequence 81, Appl
31	303.5	4.8	611	12	US-10-321-856-81	Sequence 81, Appl
32	300.5	4.8	1175	10	US-09-771-161A-224	Sequence 224, App
33	300.5	4.8	1175	10	US-09-771-161A-225	Sequence 225, App
34	300.5	4.8	1175	10	US-09-771-161A-226	Sequence 226, App
35	300.5	4.8	1212	12	US-10-247-671-157	Sequence 157, App
36	299	4.8	1233	11	US-09-291-417-89	Sequence 89, Appl
37	297	4.7	1070	15	US-10-128-714-8314	Sequence 8314, Ap
38	296	4.7	57	9	US-09-879-957-136	Sequence 136, App
39	295.5	4.7	1135	15	US-10-309-851-12	Sequence 12, Appl
40	295	4.7	1200	12	US-10-094-749-2682	Sequence 2682, Ap
41	294.5	4.7	2383	15	US-10-082-830-260	Sequence 260, App
42	294	4.7	1133	15	US-10-309-851-14	Sequence 14, Appl
43	294	4.7	1133	15	US-10-309-851-16	Sequence 16, Appl
44	293	4.7	888	11	US-09-893-519A-73	Sequence 73, Appl
45	293	4.7	929	11	US-09-298-523B-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-09-879-957-38
Sequence 38, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOMKES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEK: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 38:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 462 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 38:
 US-09-879-957-38

Query Match 32.8%; Score 2058.5; DB 9; Length 462;
 Best Local Similarity 79.3%; Pred. No. 2.6e-101; Indels 71; Gaps 1;
 Matches 391; Conservative 13; Mismatches 18; Indels 71; Gaps 1;

721 PMSTTEKPLTISAQESVYVYRALYFESRSHDEITIQGDIYMWDESQTGERGWLGG 780
 41 PGLQKKGPLTISAQENVYVYRALYFESRSHDEITIQGDIYMWDESQTGERGWLGG 100
 781 ELKGTGMPFANYAEKIPENEVPTAKVYTDLTSAAPAKLALRETPALPVTSSSEPTTP 840
 101 ELKGTGMPFANYAEKIPENEVPAVKVPTDSTSAPEKALALRETPALAVTSSSEPTTP 160
 841 NNMADFSSTWSSSNEKETDNDMTWAQPSLTVPASAGLORSAFTPATNGSSPPTL 900
 161 NNMADFSSTWPTSTNEKETDNDMTWAQPSLTVPASAGLORSAFTPATNGSSPPTL 220
 901 GQGEKEVEGLQALYPMRAKDNHLPKSDVITVLEQDDMMWFGEVQGGKMPFKSYVK 960
 221 GQGEKEVEGLQALYPMRAKDNHLPKNDVITVLEQDDMMWFGEVQGGKMPFKSYVK 280
 961 LISGVKSTSIDTQPTSPASLKVASPAKPAIPGEPIFAMTYESSBEGDLTPQCGD 1020
 281 LISGIRKSTWSDSSSPASLKVASPAKPAIPVSGE----- 318
 1021 VIVVTKDGMWTVGVGKSGVFPNSNYRLKDSGSGTAKGSLGKKPRTIAOVIAASYA 1080
 319 -----ETAAQVIASTYA 329

1081 TGPBQLTLAPGQLILIRKKNPGWMEGELQARKKRQIGMPFANYKLLSPGSKITPTE 1140
 330 TGPBQLTLAPGQLILIRKKNPGWMEGELQARKKRQIGMPFANYKLLSPGSKITPTE 389
 1141 LPKTAQVAVCVIYMTYTAQNDDELAFSGQIINVTINKEDPDWKGESVGGVGLFPSN 1200
 390 PPKSTALAAVCCVIGMYDTAQNDELAFNKGQIINVTINKEDPDWKGESVGGVGLFPSN 449

QY 1201 YVKLTMDPSQQ 1213
 DB 450 YVKLTMDPSQQ 462

RESULT 2
 US-09-879-957-194
 Sequence 194, Application US/09879957
 Patent No. US2002003475A1
 GENERAL INFORMATION:
 APPLICANT: SPARKS, Andrew B.
 HOFMAN, No. US2002003475A1h
 KAY, Brian K.
 FOWLES, Dana M.
 MCCONNELL, Stephen J.
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 USING SAME
 NUMBER OF SEQUENCES: 227
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York

COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/879,957
 FILING DATE: 13-Jun-2001
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/630,915
 FILING DATE: 03-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 1101-174
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEK: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 194:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 509 amino acids
 TYPE: amino acid
 STRANDEDNESS: <Unknown>
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 SEQUENCE DESCRIPTION: SEQ ID NO: 194:
 US-09-879-957-194

Query Match 22.1%; Score 1388.5; DB 9; Length 509;
 Best Local Similarity 48.9%; Pred. No. 8.3e-66;
 Matches 287; Conservative 85; Mismatches 124; Indels 91; Gaps 14;

635 ERKSLELEKOK---EDAQRRVQERDKOMLEHVOEQQPRPKRHEEDRLKREDSYKKE 690
 6 ERKLELMQKKLEDEARAKKQKKNLKENLRKE-----EKQKRLQEBETQKQK 59

691 AEERAKPEMDKOSRLPHQEPKATQAQWSTTEKGPLTISAQESVYVYRALYFPE 750
 60 EERKAEKQ-----RETASV-----LVNYALYFPE 86

751 SRSHDEITIQGDIYMWDESQTGERGWLGGELKGTGMPFANYAEKIP--ENEVPTPAKP 808
 87 ARNDEMSPNSGDIIVQDEKTVGEPGLYGSFQGNFGMPCNVEKMPSSENE----- 139

809 VTDLTSAAPKALALRETPALPVTSSSEPTTPNNMADFSSTWSSSNEKETDNDMTAA 868
 140 -----KAVSRKAL-----LPTVLSLATS-----TSBPLSSNQPASVYDQY--VS 180

869 QPSLTVPASAGLQRRSAFTPATATGSSSPVLGQGEKEVEGLQALYPMRAKDNHLPN 928
 181 FSNLTVNTSMQ--KKSAPFTVSPG--SVSPHGGQVVENLKAQALCSMTAKKDNHLPN 237

929 KSDVITVLEQDDMMWFGEVQGGKMPFKSYVKLISGVKSTSIDTQPTSPASLKVAS 988
 238 KHDITVLEQDDMMWFGEVQGGKMPFKSYVKIIPGSEVK-----REPEALYAAN 289

989 --PAKPAIPGEPIFAMTYESSBEGDLTFQGGDVIVTKKDGDMTGTGDKSGVPPSN 1046
 290 KKPISAAVSGEETIYPISSVSPGDLTTEGEELVLTQKDEGMWGSIGDRGIFPSN 349

1047 YVKLDSGSGTACKTSLGKKEPIAQVIAASYATGPEQLTLAPGQLILIRKKNPGWME 1106
 350 YVKEDQESFGASAKSGANKKEPIAQVTSAYVASGEQSLAPGQLILIRKKNPGWME 409

1107 GELQARKKRQIGMPFANYKLLSPGSKITPTELPKTAQVAVCVIYMTYTAQNDDE 1166
 410 GELQARKKRQIGMPFANYKLLSPGSKITPTELPKTAQVAVCVIYMTYTAQNDDE 462

QY 1167 LAFSGQIINVTINKEDPDWKGESVGGVGLFPSNYVLTMDPSQQ 1213


```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (183)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (202)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-132

```

```

Query Match      6.9%; Score 431.5; DB 11; Length 224;
Best Local Similarity 45.6%; Pred. No. 1.4e-15;
Matches 103; Conservative 31; Mismatches 51; Indels 41; Gaps 8;

```

```

QY 115 FGIGGASMPPLTAVAPVPMG-----STPVGMSRPVLSVPPAAVPLANG 161
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 FPMG---SMPNLSIPPLPPAPATSLSATSQTGLPMLMPTPLVPSVTSLSLP---NG 62
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 162 APPVIOPLPAFAHPAATWPKSSSFS--RSQPSQLMTKLQKQS--PDVASAPP----- 211
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 63 TMSLQPLP-IYSSSTPLHGSSYSLMGFG---GASIQKQSLIDLSSSSSTSTASL 118
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 212 -----AAEVAVPOSSSLKTRQLENSHDKTWSGHLTPQARTIMQSSLPQAOIASIW 263
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 119 SGNSPKGTSEWAVPQPTRLKTRQKENTLDKSMGSLGFGQARNALLQGNLSQTQLATIW 178
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 264 NUSDIDODCKLTAEFFILMHLIDVAMSGQPLPVLPPEVIPPSEFR 309
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 179 TLADXDGDQLKAEFFILMHLTMKAKQPLPLTLPPELVPSFR 224
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 8
US-09-864-761-33620
; Sequence 33620, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33620
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000193.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EST_HUMAN HIT: A473823.1, EVALUE 2.00e-37
; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-864-761-33620

```

```

Query Match      6.8%; Score 429; DB 9; Length 90;
Best Local Similarity 94.0%; Pred. No. 6.6e-16;
Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 767 VDESQTEGPGWLGLKGTGWFPPANYAEKIPENEPVTPAKPVTDLTSAPAKLARETP 826
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 1 VDESQTEGPGWLGLKGTGWFPPANYAEKIPENEPVTPAKPVTDLTSAPAKLARETP 60
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 827 APLVTSSEPTTPNNMADFSSSTW 850
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 61 APLAVTSSEPTTPNNMADFSSSTY 84
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 9
US-09-864-761-43004
; Sequence 43004, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

```

[illegible]

```
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 46565
LENGTH: 82
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUAE 9.00e-45
OTHER INFORMATION: EST_HUMAN HIT: AA773823.1, EVALUAE 2.00e-37
US-09-864-761-46565

Query Match 6.7%; Score 422; DB 9; Length 82;
Best Local Similarity 95.1%; Pred. No. 1.4e-15;
Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Db 61 APLAVTSSEPTTPNNWADFSS 82
|||||
RESULT 12
US-09-864-761-34169
Sequence 34169, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34169
LENGTH: 87
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
OTHER INFORMATION: EST_HUMAN HIT: AA815076.1, EVALUAE 4.00e-09
```

OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 1.80e-01
US-09-864-761-34169

Query Match 6.6%; Score 413; DB 9; Length 87;
Best Local Similarity 96.6%; Pred. No. 4.5e-15;
Matches 84; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 348 TPEKKEKNEFGRGVELEKROALLFOORKEOEELAQLEAEORERKEROEAKROLE 407
DB 1 TPEKKEKNEFGRGVELEKROALLFOORKEOEELAQLEAEORERKEROEAKROLE 60

QY 408 LEKOLEKORELEKROEERERKEIRRE 434
DB 61 LEKOLEKORELEKROEERERKEIRRE 87

RESULT 13
US-09-764-881-128
; Sequence 128, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT207
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 128
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (122)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (155)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (147)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (165)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-128

Query Match 6.5%; Score 409.5; DB 11; Length 168;
Best Local Similarity 60.4%; Pred. No. 1.5e-14;
Matches 87; Conservative 18; Mismatches 32; Indels 7; Gaps 4;

QY 1 MAOPPTFGSLDVAIVVERAKDQFLSLKPIAGITDQANFFQSGLPQVLAQ 60
DB 8 MAOPPTFGSLDVAIVVERAKDQFLSLKPIAGITDQANFFQSGLPQVLAQ 67
QY 61 IMAADANNDGRMPDQVEFSIAMKLIKLIKLGQYQLPSTLPVWKQOPV--AASSAPAFGIG 118
DB 68 IMAADANNDGRMPDQVEFSIAMKLIKLIKLGQYQLPSTLPVWKQOPV--AASSAPAFGIG 126
QY 119 GIASMPITJAVAPVPMGSIPIVGM 142
DB 127 --SMPNLSIPQPLPPAA-PITSL 146

RESULT 14
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US2003013948A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Collier, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/10/294,804
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/410,399
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

Query Match 6.1%; Score 382; DB 12; Length 1162;
Best Local Similarity 24.6%; Pred. No. 4e-12;
Matches 127; Conservative 139; Mismatches 197; Indels 54; Gaps 19;

QY 327 QRLPE--EPSEDEBQ--PEKULPTVEDKKNFGRGVELEKROALLFOORKEOERL 382
DB 551 QRSFQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQREPPQ 608
QY 383 AOLERAEOERKEROEAKROLELEKOLEKORELEKROEERERKEIRREAKRELER 442
DB 609 EQ--QDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDE 664
QY 443 QROLEWERNRROELNQRNKEQEGTVLTKARRTLFELEBALNDKQKQLEGLQDIRCRL 502
DB 665 QQQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQ 709
QY 503 ATQROEISTFNKRELIATITLQOOLQESQMLGRLFEKQILSLQKQVQNSLHRD 562
DB 710 DEBQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDEQDE 767
QY 563 SLTLKRALAKELARQLEQLDEVERETRSLQLEIDVNNOLKELREIHSKQOLQKO 621
DB 768 ELEBQDELEDBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQ 822
QY 622 RSLR--AARLKQKOEKRSLELEKQEDAQRRVOERDKQMLEHVQDEQPRPRKPHR 678
DB 823 QLEBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDE 882
QY 679 RLKREDSVRKKEAEERAKPEMCD---KQSRLEPHHOEPKATQAPWSTTEKPLT 732
DB 883 EYEQEQLBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDELEBQDE 941
QY 733 -SAQESVYVYVYALYFESRSHDEI-TIQGDIWVDESOTGPGMLGSELKGTGMP 790
DB 942 SSSSEDEVD---YVVS--THEQIASPPGDTTPDDDPGPGSRERYVLRKISPPHR 995
QY 791 ANYAEKIPENEVPTPAKPVDTLSAPAPKLALRETPA 827
DB 996 GVRMRKVP---VTHPKKPHRYQGPVPVYRQIDDCPA 1029

RESULT 15
US-10-171-311-83
; Sequence 83, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John

```

; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatc, Karen
; APPLICANT: Ganavara, Manjula
; APPLICANT: Hoerh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PaateSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 2701
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-83

Query Match      5.8%; Score 363.5; DB 15; Length 2701;
Best Local Similarity 20.7%; Pred. No. 1e-10;
Matches 280; Conservative 165; Mismatches 490; Indels 415; Gaps 59;

QY 20 ERBAKDOF--LSLKP--IAGFTGDQARNFFFGSLPQVLAQIMALADNNNDGRMD 74
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 155 KKEKTDNDYGPSPSLRPPVACWGRGAAAGSSSDQDEKLPQO-----BESTAGTSE 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 75 Q-----VESTAMKLIKIKLOGYOLP-----STLPVWKOQPAVISAAPAFIGIGIAS 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 210 QNDILKVKRRIACGPPQAKNGQALASQYRAMMPYTFQYPRMTYPLPHG----- 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 MPELTAVALPVMGSLPVVGMSPPLVSSVPPAAPPVPLANGAPVYIOLPLPAFAHPAATWPKS 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 264 -----PM-----RPPSLSETKKG-----LRGGRP-----PSM-----ASEPFR 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 183 SPSFSGSPGSQLNTKIQKAKOSPDVASAPPAEMAVPOSSRLKTYROLFNSHDKTMSGHLTG 242
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 294 PSLISA-----SELKEIDKFNDLDAEADGWAQAQ-MEVDYEQLNFSDDDEQGS-NS 344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 243 PQARTILMOSLPQAQALASIMNLSIDIDQOKLJAEFIILAMHLIDVAMGQPLPVLPE 302
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 PKENNSDQGS-----KASENNENKKTETDE-----VSNTKSSSQITP----- 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 303 YIPSPFRVRSQSGMSVSISSSVQRLPREP-----S 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 381 -AOPSVAKYPIYKGPSPFNQERGTSSHLPPPKLLAQOHPPRPOAVPGRPGPPSKQVA 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 335 SEDE-----QQPEKLLPVTEDEKRENFERSGVELERQALLBQQRKEQERLAQLER 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 440 DEDEIKQRRROQSEISAVERARKRREBER-RME-EQRKAAQAKLRLDEKLGITLX 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 --AEQKREKREQEOAKQOLEK-----QLEKQELERQEBERKEITERREAAKRELE 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 498 QPSPEIRERER-EKEREKELEKEQEQEREKEREKREKREKQEKKELEKEQEKREME 556
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 442 RQR-----QLEWERNRROELNQRNKEQESTVVLKARKTLEFELALNDKQJLEGLQD 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 557 KERKQEKELERQEKELERQEKELERQEKELERQEKELERQEKELERQEKELERQEKELER 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 498 INCRILATOROE-----TESTNKSRELRLIAETIH----- 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 614 QSENSCNKEEBEVFTRODSNRS-EKEATPVVHETEPESGQPRPAVLGTFKQOKSLP 672
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 -----LQOQLOESOO----- 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 673 PRFORQOEOMKQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 732
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 536 -----MLGR-----LIPKQILSDQLKQVQONSILH----- 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

DB 733 FDBRMLMOSYMDPRMMSGRPAMDIPRIHGMIPPKPLMRRRQMGSSPSSSEFHIARS 792
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 561 -RDSLLTLTKRA-----LEAKEIARQOLREQLDEVERETMSKIOET---DVFN 604
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 793 ARDHAISISEPRLMGSDPYPHAEPOQATTPRATSEPEDVRSEALDQOQITAAVSVEHN 852
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 605 OKE-----LREHISKQOLQK--QRSLEAALTKQEKQERKSLLEKQEKEDQRRQVED 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 853 QLEAHPKADFIRE-SSEAQVQKFLRSVEDVPPHHTDANNQACFAPDQKTLAPQE-- 909
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 657 KOMLEHVQOEQPRPRK--PHEEDRLKREDSVR-----KXEAERAPX 697
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 910 -----ERISAVESQPRKRSVSHGSHHTQKPDQGRBPSAGIPKYSRCLDSKEPIER--P 963
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 698 EMQDKOSRLFHPHQBPAKIATQAPWSTTEKPLTISAQSVKVVYRALLYPRESSHDEI 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 964 EEKPKKEGFIRESSEGPKEKVKYSKSETWGRP-----SSNRREBV 1005
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 758 TIQP-----GDI-----VMVDESQTEGPGWLQGLKGTQGWMPANYAEKIPENEVPTPAK-PV 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1006 NDRPVRSGPIKKPVLKRDKEERE-----QKREKQ-----EKAERKVEKVVAKDEKTEK 1055
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 TDLTSAPAPKALARETPAPLPYTSSEPSSTPPNNMADFSSSTWPDSSSNEKETDNMDTMAAQ 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1056 KDLPPPPPPP-----QPPAPIQPSVPPPIQPEBA-EKPFETETATLAQKRSQDT--EKPLE 1108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 870 PSLTV--PSAGQLRQSAFTPATATGSSPSVJLQGEKVEGLQAOALYPMRAKDNHNLN 926
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1109 PVSTQVEBPAVTVNQATTAAPVAVKEKQPEKVISKDLVIE-----RPRDSRPA 1158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 927 FNKSDVITLBEQDDMMNGEVOGQKWP-----KSYVKLISGVPKST 970
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1159 VKKE--STLPPRTYW---KEARERDWPFDQYRGRGEGEYYSRGRSTYSGYGRGKG- 1211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 971 SIDGPTESPASLKEVASPAKPAIPGEEFIAMTYESESQGLTFQOQGVIVTVTKODG 1030
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1212 ---RHTHD-----YPOYDNKPRAEHIPSGLRQRESETSSESDFEVVKKRRQ 1260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1031 WMTGVGDSGVFPBNYVRLKDSGSGTAGTKTSGLKKPELAIQVIASTAAATGPEQULTAP 1090
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1261 RQSETDITD-----SEIHESASDKDSLKSKGLPKREERPEKPKVPKPHSSFKPDNHVRID 1314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1091 GOLL---LIR---KKNPGMWEGELQARGK 1114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1315 NRLLEKPYVRDDKAKPGFLPKGEPTRRGR 1344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: December 4, 2003, 15:25:42
Job time : 36 secs

THIS PAGE BLANK (USPTO)